E261 530

AD-A265 034

PL-TR-92-2344

DATABASE BLENDING FOR THE CLIMATOLOGY OF CLOUD STATISTICS PROGRAM

James H. Willand

Hughes STX Corporation
109 Massachusetts Avenue
Lexington, MA 02173

26 December 1992

Final Report
26 April 1988-25 November 1992

Approved for public release; distribution unlimited

PHILLIPS LABORATORY
Directorate of Geophysics
AIR FORCE MATERIEL COMMAND
HANSCOM AIR FORCE BASE, MA 01731-3010

"This technical report has been reviewed and is approved for publication"

Donald D Grantham

DONALD D. GRANTHAM, Contract Manager
Chief, Atmospheric Structure Branch

Robert A. McClatchey

ROBERT A. McCLATCHEY, Director
Atmospheric Sciences Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	26 February 1993	Final Apr. 26, 1988 - Nov. 25, 1992

4. TITLE AND SUBTITLE

DATABASE BLENDING FOR THE CLIMATOLOGY OF CLOUD STATISTICS PROGRAM

5. FUNDING NUMBERS

F19628-88-C-0089
PE 62101F
PR 6670
TA 09
WU AP

6. AUTHOR(S)

James H. Willand

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Hughes STX Corporation
109 Massachusetts Avenue
Lexington, MA 02173

8. PERFORMING ORGANIZATION REPORT NUMBER

Hughes STX Corporation

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Phillips Laboratory
29 Randolph Road
Hanscom AFB, MA 01731-3010
Contract Manager: Donald Grantham/GPAA

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

PL-TR-92-2344

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report describes data sources, data processing activities, compression, prediction, and blending of the climatological databases chosen to represent realistic global cloud amount statistics within the C Cloud S program.

14. SUBJECT TERMS

Clouds, Climatology, NIMBUS-7, Coverage

15. NUMBER OF PAGES

66

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited Distribution

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF TABLES

ACKNOWLEDGEMENTS

This work was supported by the Phillips Laboratory, Geophysics Division, Contract Number F19628-88-C-0089. Technical assistance from Mr. A. Boehm of Hughes STX is gratefully acknowledged.

1. INTRODUCTION

This report describes the data sources, the data processing activities (data extraction, organization, compression, and prediction), and the blending of sky/cloud cover statistics which are vital to the development of an automated fast retrieval global Climatology of Cloud Statistics (C Cloud S). The material presented supersedes the work presented in an interim report by Willand and Steeves (1991). It begins with a brief description of the overall C Cloud S system. The sections that follow describe in detail the processing and blending of the climatological databases chosen to represent realistic global cloud amount statistics within an automated C Cloud S program.

2. C CLOUDS S SYSTEM OVERVIEW

C Cloud S is based on many years of cloud observations. It does not provide observed values for a given date and time in the past and therefore is not a historical database. Instead it is intended to be used directly as either input to simulation models or to provide a synthetic series of cloud or cloud-free line-of-sight predictions that have realistic probability distributions and are realistically correlated in time and space.

An overview of the complete C Cloud S system is portrayed in Figure 1. As shown, the system is divided into three subdivisions: Data Input, Compact Storage, and Cloud Statistics plus or minus Error.

2.1 DATA INPUT

Data input to the C Cloud S system consists of four climatological data ensembles

1. The Burger cloud climatology from land observations
2. Department of Energy (DOE) cloud climatology from land and sea surface cloud observations
3. The NIMBUS-7 satellite CMATRIX cloud climatology and
4. A subset of the DATSAV climatology that provided hourly sequences of surface sky cover observations used for deriving temporal and spatial correlations of cloud covers.

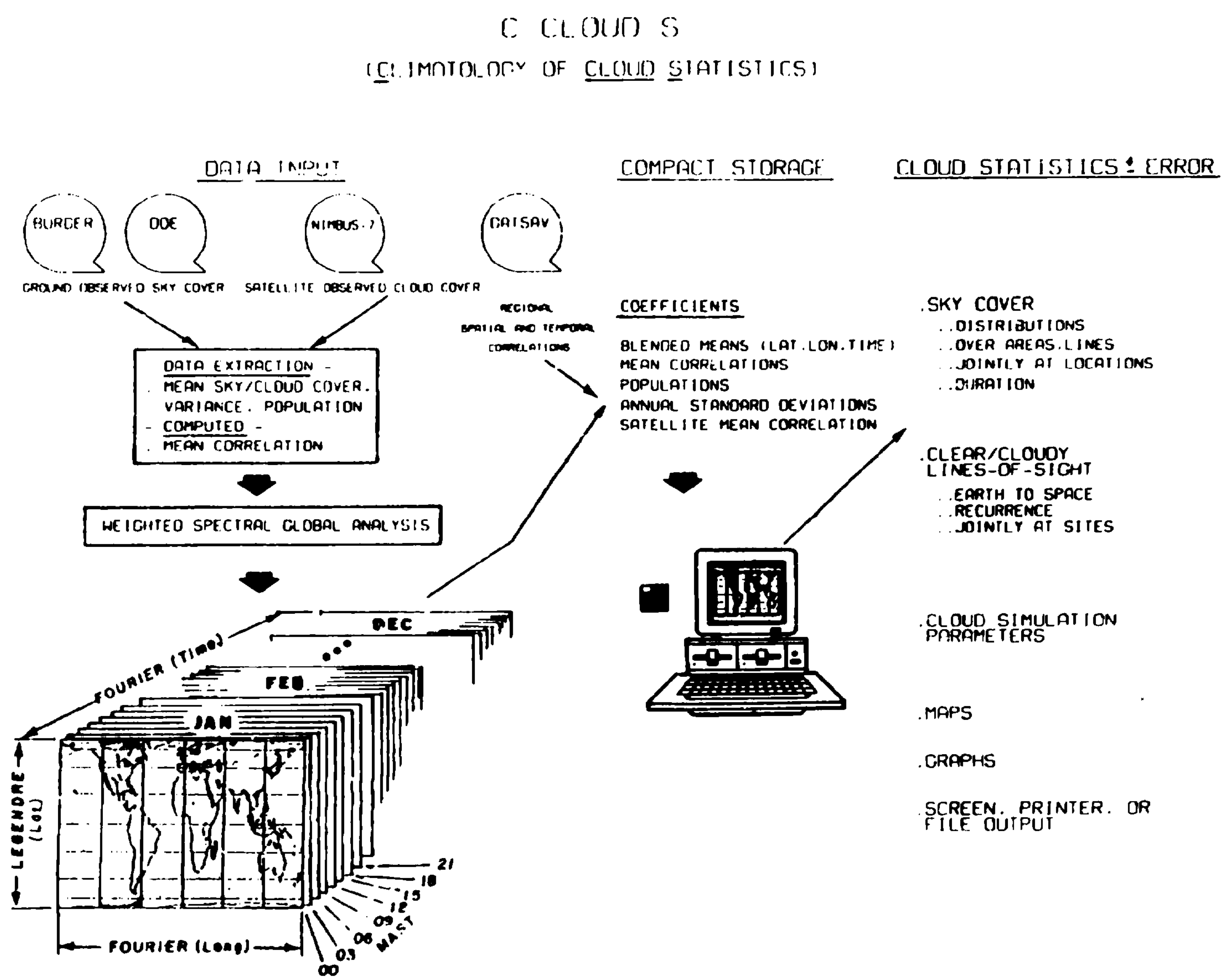


Figure 1. C Cloud S System Overview from Start to Finish.

Spatial and temporal attributes of these climatic databases are summarized in Table 1. Note that none of the spatial or temporal attributes of the databases are the same. For example, DOE/NCAR data are mapped into a 5 x 5 degree grid system, whereas NIMBUS-7 data are in a 2070 Earth Radiation Budget target area grid. Burger data and DATSAV data are sky dome point climatologies gridded randomly. DOE data are stratified by season at every three hours GMT while NIMBUS-7 data are archived daily at noon and midnight MAST (mean apparent solar time). Burger data, on the other hand, are mid-seasonal monthly means and histograms of three hourly mean sky covers centered over four local standard times of day. DATSAV data are mostly archived 24 hours daily in GMT. Finally, Burger, DOE, and DATSAV data archive sky cover amounts that are estimated by surface weather observers while the NIMBUS-7 data archives cloud amounts as observed from space. (Sky cover is defined as the amount of a sky dome area covered by clouds as estimated within the perspective of an up-looking device or a ground observer. Cloud cover on the other hand is the amount of earth covered by clouds within a given field of view as observed from the down-looking perspective of an earth viewing satellite.)

Table 1. Summary Table of Cloud Databases for C Cloud S 1.0

Database	DOE/NCAR	NIMBUS-7	Burger	DATSAV
Source	Surface Observed sky covers within an area.	Sun synchronous satellite derived cloud cover.	Surface Observed sky covers over stations.	Surface observed sky covers over stations.
Area Resolution	555 x 555 km ~ (5 x 5) deg	500 x 500 km	sky dome	sky dome
Points/Map	1820	2070	2300	92
Yearly Stratified	4 seasons	Daily	Mid-seasonal monthly.	Hourly
Times of Day	3 hourly 0,3,...GMT.	Twice/day, 00–12 MAST.	4 hours LST 1, 7, 13, 19.	24 hours GMT.
Period of Record	Land, 11 years, 1971–1981. Ocean, 54 years, 1930–1983	6 years, 1979–1985.	29 years, 1945–1973.	11 years, 1973–1983 or 10 years, 1977–1986

2.1.1 Data Extraction

Data extraction methods were initiated to extract from these climatological data ensembles the mean sky/cloud cover populations and their variances and interannual standard deviations. Computed mean correlations, defined below, were also assembled using mean sky/cloud covers and variances extracted for each of the points encountered within a dataset. (Parameterization of sky cover distributions over selected locations can be generated by combining the two parameters of mean sky cover and mean correlation. The algorithm for the parameterization can be found in Boehm (1992).) All extracted and computed parameters including date-time groups and latitude-longitude locators were stored in a convenient standard format for subsequent processing.

Mean correlation is defined as the mean of the correlations of given conditions (such as sky/cloud coverage) between all pairs of points in a domain where a domain can be either a line, an area, or a volume.

2.1.2 Weighted Spectral Analysis

A weighted spectral analysis program was developed and utilized to compress the extracted data into the form of spatial coefficients using a Fourier analysis in longitude and a fully normalized associated Legendre function in latitude. The coefficients derived are used to predict five basic parameters given the coordinates of any location on earth. These five parameters are mean sky cover, mean sky dome correlation, population, interannual standard deviation of mean sky cover, and satellite mean correlation. A Fourier series can be used to predict cloud amount statistics for any time of year and time of day.

2.2 COMPACT STORAGE

The number of sets of compact storage coefficients for each of the five selected parameters are listed below.

1. Blended mean sky/cloud cover statistics from Burger, DOE, and NIMBUS-7 data. (12 months x 8 times or 96 sets)
2. Mean Correlations over a sky dome as computed from Burger and DOE data. (12 months x 8 times or 96 sets)

3. Populations derived from blending Burger, DOE, and NIMBUS-7 data sample sizes. (12 months x 8 times or 96 sets.)
4. Annual standard deviations of sky cover statistics derived from DOE data. (12 months or 12 sets.)
5. Satellite mean correlations derived from NIMBUS-7 data. (12 months x 2 times or 24 sets.)

The sets of coefficients were spooled onto floppy disks for use in the development of an automated cloud climatology to be housed on small personal computers. Regional spatial and temporal correlations of sky cover conditions over selected DATSAV data stations were plotted up and used to define equations for predicting joint cloud amount events within the automated cloud climatology program.

2.3 CLOUD STATISTICS PLUS OR MINUS ERROR

In summary, C Cloud S (Climatology of Cloud Statistics) is a compact global cloud climatology residing on an IBM compatible computer. The basic databases are derived from NIMBUS-7 satellite data, DOE/NCAR surface data, and DoD surface data. Cloud models based on special datasets are used to calculate derived statistics. The following statistics together with estimated error are available for any point on the globe, any time of year, any time of day:

SKY COVER STATISTICS

1. Probability of Fractional Sky Cover
2. Duration of Sky Cover Category
3. Recurrence of Sky Cover Category
4. Conditional Climatology of Sky Cover.

CLOUD-FREE LINE-OF-SIGHT

5. Probability of CFLOS
6. Duration of CFLOS
7. Recurrence of CFLOS
8. Joint Probability of CFLOS at N Sites

SATELLITE-BASED VIEWING

9. Cloud Coverage Over an Area
10. Cloud Coverage Over a Line
11. Probability of LOOK/SEE

DATA BASE PARAMETERS

12. Mean Sk Cover
13. Sky Do Mean Correlation
14. Large rea Mean Correlation
15. Effective Period of Record

Output is in the form of numbers, graphs, histograms, or maps. Output can be sent to the screen, to a printer, or to a file for use by other computer programs. Limitations: C Cloud S is limited to total sky cover, not clouds at given altitudes.

2.4 CLIMATOLOGICAL DATABASES AND DATA PROCESSING ACTIVITIES

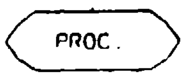
The following section and sections 3 through 7 describe the climatological databases used for input to C Cloud S. Also, data processing activities such as data extraction, organization, interpolation, analysis, compression, and blending of these data ensembles are described in detail. Flow diagrams are included to demonstrate how data sources are combined with developed software to produce end results. These diagrams all use a common form of symbols defined in Table 2. The diagrams show data sources used by data transfer software developed to format, compress, and transfer data to permanent files. Backup tapes and labels record where all data for each project are kept for permanent retention. Most of the data processing activities subsequently described were performed on a CDC 6800 CYBER NOS system or a CONVEX UNIX system. The software was written in standard FORTRAN 5. Several FORTRAN subroutines and functions critical to the processing effort are listed in Appendix A.

3. BURGER DATA

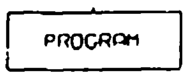
The Burger data (Burger, 1985) were compiled mainly from the Revised Uniform Summaries of Surface Weather Observations (RUSSWO's), National Intelligence Surveys (NIS), and NAVATLAS records. The resulting data ensemble contains station names, latitude-longitude locators, fractional sky cover histograms,

Table 2. Symbol Definitions Used in Data Processing Configuration Diagrams

Magnetic Tape Data Storage

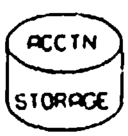


Procedure Used for Batch Processing



Computer Program (FORTRAN)

Resulting Product or Other Programs



Account Number

Permanent Disk Data Storage

Floppy Disk Storage

mean sky covers, scale distances (Burger, 1985), and in some cases, period of record for over 2,000 weather reporting stations around the globe. The data are archived for the four mid-season months of January, April, July, and October. Three hourly averages of sky cover are centered at the four local standard times of 01, 07, 13, and 19 LST. Data over the ocean areas are archived for the same four mid-season months, but sky cover values are monthly means. Willand (1988) revised the original digitized version of the Burger sky cover archive of RUSSWO's, NIS, and NAVATLAS data into a more compact compatible format. The data are now stored in three separate files called RUSARC, FINARC, and SEAARC, respectively. Subsequently the revised dataset was used in the development of a prototype automated global cloud climatology (CLOUDZ), Boehm and Willand (1988). A detailed description of the processing of the data required for the development of the latest version of the automated climatology of cloud statistics is discussed next.

3.1 PROCESSING OF BURGER DATA

The data flow configuration for processing Burger data is shown in Figure 2. The FINARC data portion of the Burger data were mainly utilized in the following processing efforts. Except for a few selected southern hemisphere stations, the RUSARC and SEAARC portions of the Burger data were excluded from the processing because of possible overlap with the DOE climatology (discussed in Section 4.0).

3.1.1 Program SDTORO

Initially, FINARC data were spooled from backup tape M14555 on to permanent disk file. The procedure XEQSDT was then used to call on program SDTORO which read the data and converted the scale distance parameters to mean correlations (see subroutine SKYRO, Appendix A). Missing sample sizes for stations showing climatological sky cover values were set to 450 observations, or 5 years x 30 days x 3 hours. The resulting data together with date-time information and locators were written to permanent disk files in a standard format designed to be utilized in interfacing with other datasets. Thus, a date time group, latitude-longitude, mean sky cover, standard deviation, mean correlation, and population count for each station for each separate month and time period were written to permanent disk storage using the following format:

```
    WRITE(NTAPE,100)ITC,ISEA,ITM,ITIM,LAT,LON,M,S,RO,P
100 FORMAT(I3,I2,I1,I2,F6.2,F7.2,3F6.3,I8)          (3.1.1)
```

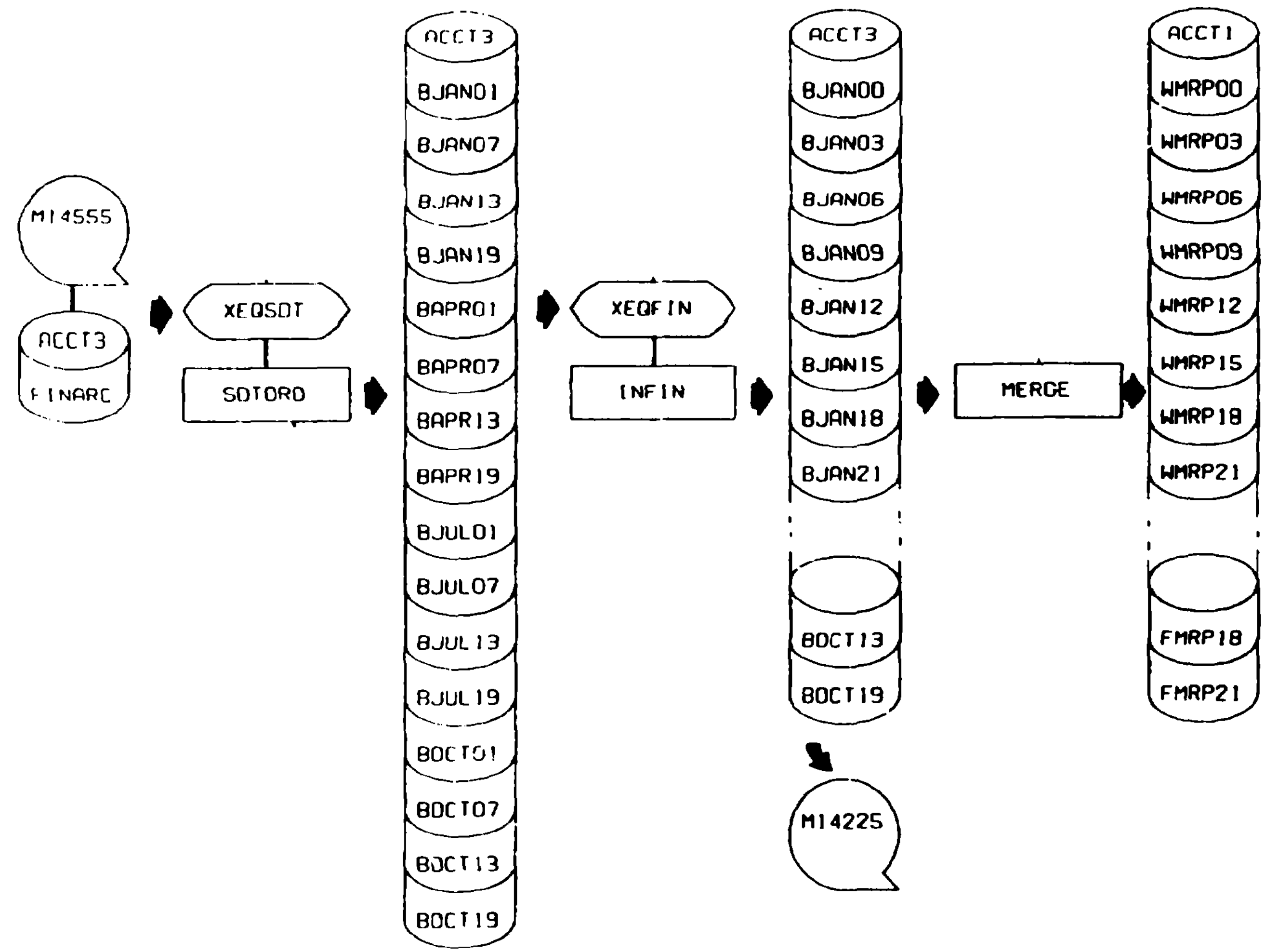


Figure 2. Flow Diagram for Processing Burger FINARC Sky Cover Data into a Form Suitable for Merging with Other Data Files.

Variables are defined as

ITC:	Data type code.
ISEA:	Month/Season code.
ITM:	Time Code.
ITIM:	Time. 00,03,06,... 21.
LAT:	Latitude. -90 <= LAT <= 90 (REAL)
LON:	Longitude. 0 to 360 degrees east. (REAL)
M:	Mean cloud cover probability. 0.0 <= M <= 1.0 (REAL)
S:	Standard deviation. 0.0 <= S <= 1.0
RO:	Mean correlation. 0.0 <= RO <= 1.0
P:	Population

The converted data were written to account number 3 disk storage under names like BJAN01, 07, 13, 19 and BAPR01, etc., meaning Burger data for January, April, July, or October at 01, 07, 13, or 19 LST respectively.

3.1.2 Program INFIN

INFIN uses linear interpolation methods to interpolate the data written by SDTORO to 8 times of day beginning with 00 LST. Thus, resulting data was again written to permanent files only this time with names like BJAN00, 03, 06, ...21LST or BAPR00, 03, 06, etc. The final datasets were spooled to tape M14225 for permanent retention. The INFIN process was necessary in order to obtain data in a form suitable for merging with the DOE data discussed in Section 4.0.

4. DOE/NCAR DATA

The "Climatological Data for Clouds Over the Globe from Surface Observations" compiled by Hahn (1987) was another important sky cover database used in the development of C Cloud S. This cloud database was prepared and archived for use in several U.S. Government energy agencies, including the United States Department of Energy (DOE) and the National Center for Atmospheric Research (NCAR). The database contains global maps (90 deg N to 90 deg S) of long-term monthly and/or seasonal total sky cover, cloud type amounts and frequencies of occurrence, low cloud base heights, harmonic analysis of

annual and diurnal cycles, interannual variations and trends, and cloud type co-occurrences. Most of the data are mapped on grids. Each grid box represents a 5-degree latitude by 5-degree longitude resolution equatorward of 50 degrees latitude. Poleward of 50 degrees, the longitudinal length of the boxes is increased to maintain approximate equal area in each box. According to Hahn (1987), data from all available stations within a grid box were used to compute the average for that box. The data were archived so that land-observed sky cover data were stored separately from ocean-observed data. The period of record for land observations is the 11 years from January 1971 through December 1981. The period of record for ship observations is the 54 years from January 1930 through November 1983. Portions of the DOE data archive are available in the form of atlases (Warren et al., 1986), and considerable additional information is available on a single magnetic tape.

4.1 PROCESSING OF DOE DATA

The computerized system for processing the DOE data is depicted in Figure 3. The portion of DOE data selected to represent sky cover statistics within the C Cloud S system was that archived at 8 times of day and averaged over the four seasons Winter, Spring, Summer, and Fall. These statistics provided valuable information for defining and predicting diurnal cycles of global sky cover.

4.1.1 Program FETCH

Program FETCH, which is invoked by the procedure XEQFET, extracted global maps of selected cloud statistics from separate files of the DOE cloud climatology tape. The files had been assigned map group numbers specified by Hahn (1987). Specifically, selected global maps of mean sky cover for four seasons at eight GMT's daily over land areas were extracted from file 2 under map group numbers 14 through 45. Ocean data for the same parameter were extracted from file 6 under map group numbers 1080 through 1111. The extracted data were stored under the names NC1080, NC1096, etc. The parameters stored for each box encountered were mean sky covers and standard deviations, populations, central box latitude-longitude reference points, and date-time information.

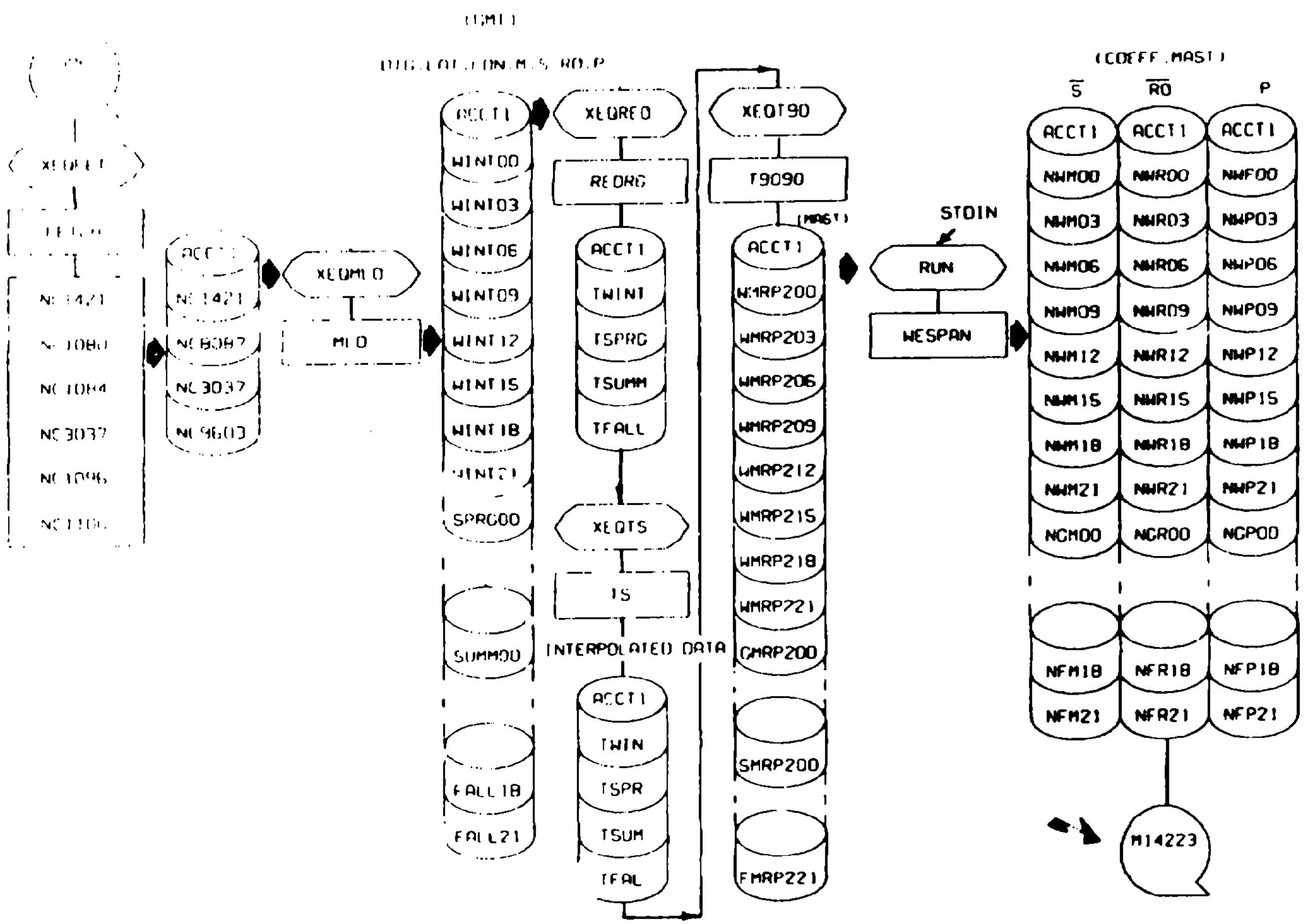


Figure 3. Data Flow Configuration Used in Processing the DOE Seasonal Cloud Databases.

4.1.2 Program MLO

Procedure XEQMLO calls on program MLO in order to merge DOE land observations of mean sky cover with ocean observed mean sky conditions. During the merging process, the standard deviation of mean sky cover for each 5-degree x 5-degree grid box was adjusted by

$$s=\frac{s'(n-1)}{n} \tag{1}$$

where

s´= DOE standard deviation of mean sky cover at a grid point

n = sample size

s = adjusted standard deviation.

In addition, mean correlation was computed using the given mean sky cover amount and the adjusted standard deviation squared for each grid point as input to an accurate tetrachoric correlation function like that listed in Appendix A and described in Smyth (1991). Normally the four arguments A, B, C, and D sent to this function represent normalized counts in a fourfold contingency table of dichotomous events. However, in order to compute the mean correlation for each box encountered these arguments were set to the following computed values.

$$A = V + M \times M \tag{2}$$

$$B = M - A \tag{3}$$

$$C = B \tag{4}$$

$$D = 1.0 - A - 2C \tag{5}$$

where

M = Mean sky cover

V = Variance

then

Mean Correlation = TETRA(A,B,C,D).

A date time group, latitude-longitude, mean sky cover, standard deviation, mean correlation, and population count for each grid point for each separate season and time period were written to permanent disk storage with the standard format shown in 3.1.1. Names such as WINT00 (winter at 00 GMT) were assigned to the data files. A total of eight files at eight GMT's for each of four seasons were assembled.

4.1.3 Program REORG

Program REORG initiated by procedure XEQREO was implemented to reorganize the datasets created by program MLO and thus prepare them for subsequent interpolation. The 8 files for each separate GMT for each season were extracted, reordered and stored in the format 4.1 below.

```
      WRITE(NTAPE,400)YLAT,XLON,(XM(IT),IT=1,8),(XR(IT),IT=1,8),
     X                         (IP(IT),IT=1,8)
  400 FORMAT(1X,F5.1,F6.1,2(8F5.3),8I5)              (4.1)
```

where

YLAT = Latitude

XLON = Longitude

XM = Mean sky cover

XR = Mean correlation

IP = Population

IT = Time index 1 through 8.

The reordered datasets were named TWINT, TSPRG, TSUMM, and TFALL meaning winter, spring, summer, and fall respectively. These newly formed data values were then better suited for use in the time series interpolation process described next.

4.1.4 Program TS

Procedure XEQTS calls on program TS (Time Series interpolation) to linearly interpolate missing mean sky cover, mean correlation, and population values within the reorganized DOE dataset. For instance, many of the DOE data values were archived every four hours GMT (e.g., 00, 06, 12, and 18Z), but often data for the in between times (e.g., 03, 09, 15, and 21Z) were missing. The interpolation procedure to fill in these missing data values provided a more stable data set for use with a more robust Fourier

interpolation scheme (Section 4.1.5) used to convert the entire data ensemble in GMT to mean apparent sun time, MAST. The final TS interpolated data ensembles were again stored on separate files for each of the four seasons. The files were given names such as TWIN, TSPR, TSUM, and TFAL.

4.1.5 Program T9090

The data values within the DOE dataset at this point of the data processing activity were all representative of Greenwich Mean Time (GMT). In the development of an automated cloud climatology it is more desirable to have the values representative of Mean Apparent Solar Time (MAST). Therefore, program T9090 was developed to transform the DOE data from GMT to MAST. The data ensembles created by program TS were used as input. The T9090 program then utilized a Fourier series in time to transform the data from GMT to MAST.

The procedure begins by sending each 8 hour set of values for a given parameter, (e.g., mean sky cover, mean correlation, or population for a given station within a given season) through an analysis routine that uses Cholesky regression initially and then a Fourier series to create 8 coefficients. These coefficients were then used to predict the value of a given parameter for any given time (GMT). The GMT used for prediction at a given point and MAST was computed using 6.

$$GMT = -24 \text{ Hrs/360 Deg} \times ELONG + MAST \tag{6}$$

where

ELONG = Longitude (Degrees East)

Tests were then made to ensure that GMT remained within a 24-hour cycle. Tests were also made on the magnitudes of predicted population values in order to keep them within tolerable limits for numerical stability. The maximum predicted population value set for the winter season was 990 observations, spring and summer was 1012, and fall was 1001.

This process was repeated for every point in a DOE grid system for each parameter creating a synthesized data file at every three hours MAST. These data files consisted of a date time group (DTG), latitude-longitude locators, mean sky cover (M), standard deviation, mean correlation (R), and population (P) for winter (W), spring (G), summer (S), and fall (F) in the standard format shown in 3.1.1. Names such as WMRP200 were assigned to each file. Here, 2 represents the version number and 00 the MAST.

Note that at this point, the interpolated Burger FINARC data discussed in Section 3 were merged with the resulting MAST version of the DOE data.

4.1.6 Program WESPAN

The merged DOE/BURGER datasets in MAST were transferred to a UNIX computer system for subsequent processing through an analysis program called WESPAN, WEighted SPectral ANalysis. The program was initiated by a Bourne shell script called "RUN" that resides on "STDIN".

4.1.6.1 WESPAN Weighted Spectral Analysis

Spectral analysis is accomplished by using a weighted spectral analysis in longitude with fully normalized associated Legendre functions in latitude. Since data is irregularly spaced (and weighted), the weighted correlation matrix is inverted rather than separately calculating coefficients which assume orthogonality between the spectral analysis functions. Note that weights are not applied when doing the analysis of population values. Values of mean sky/cloud cover and mean correlation are converted to equivalent normal deviates before performing the calculations described below.

The variance of the mean value is proportional to the number of cases regardless of the time correlation structure. Thus the weights are:

$$w_i = N_i \tag{7}$$

where w_i is the weight at the ith location and N_i is the number of days of data at the same location. N_i in this case is not allowed to be greater than the population limits set in Section 4.1.5. These limits are well within bounds since climate studies show that the variance of the mean does not grow much smaller after about 15 years of data for a given month (15 years x 31 days = 465).

The analysis described here, except for the inclusion of weights, is much like any other curve fitting technique where lots of data are compressed into coefficients which are subsequently used to predict certain parameters within some degree of accuracy at a given state. Thus given that **L** is the matrix of locations and N is the number of locations:

$$
L = \begin{bmatrix} lat_1, & lon_1 \\ lat_2, & lon_2 \\ & \vdots & \\ lat_N, & lon_N \end{bmatrix} \tag{8}
$$

and **V** is the vector of values (e.g., mean sky cover, correlation or population) at each point

$$
V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} \tag{9}
$$

and finally **W** is the diagonal matrix of weights, i.e. $w_{i,i}$ is the weight for the ith location,

$$
W = \begin{bmatrix} w_{1,1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{N,N} \end{bmatrix} \tag{10}
$$

it is desired to derive a regression fit to the data of the form:

$$
v = \sum_{i=0}^{224} a_i S_i(lat,\ lon) \tag{11}
$$

where a_i are regression coefficients and S_i are spectral analysis functions of latitude and longitude, e.g.,

$$
F \cos(2\ lon) \sin(lat) \tag{12}
$$

where F is a normalization constant different for each spectral function. To solve for the values of a_i, we first transform the location matrix **L** into a matrix **S** that gives values of spectral analysis functions at each location.

$$S = \begin{bmatrix} s_{1,1} & \cdots & s_{1,m} \\ \vdots & \ddots & \vdots \\ s_{N,1} & \cdots & s_{N,m} \end{bmatrix} \tag{13}$$

where m is the number of spectral analysis functions, or 225 in this case. In the program these are labeled 0 to 224.

The mean value and standard deviation of each spectral function is calculated using the value of the function at each of the N locations and weight at that location. If the locations were uniformly spaced over the globe and the weights were all one, then the mean value would be zero.[1] Using these mean values and standard deviations, the spectral analysis functions are normalized by subtracting the mean and dividing by the standard deviation. These normalized values are identified by use of a caret over the variable in equations 14 and 15 below. Thus the weighted correlation matrix **R**

$$\boldsymbol{R} = \hat{\boldsymbol{S}}^T \boldsymbol{W} \hat{\boldsymbol{S}} \tag{14}$$

and the vector $\hat{\boldsymbol{a}}$ of normalized weighted regression coefficients is found using (Tompkins and Wilson, 1969):

$$\hat{\boldsymbol{a}} = \boldsymbol{R}^{-1} \hat{\boldsymbol{S}}^T \boldsymbol{W} \hat{\boldsymbol{V}} \tag{15}$$

The actual inversion is accomplished using the Cholesky square root algorithm (Westlake, 1968). Then the normalized regression coefficients, $\hat{\boldsymbol{a}}$, are converted to ordinary regression coefficients by multiplying

[1]This is one major difference from the typical use of spectral functions in numerical weather prediction.

by the standard deviation of v and dividing by the standard deviation of the corresponding spectral function. Thus

$$a_i = \frac{\hat{a}_i \sigma_v}{\sigma_{s_i}} \qquad (16)$$

Eight sets of coefficients for each of the 8 MAST's for each season for the 3 parameters mean sky cover (S), mean correlation (R()), and population (P) were stored on permanent files with names like NWM(X) that stands for New Winter Mean coefficients for (X) MAST. These sets of coefficients and accompanying software were stored on tape M14223 for backup purposes. The coefficients were written out using the format below:

```
      WRITE (NTAPE,100) I, A(I)
  100 FORMAT(I6,E15.7)                              (4.1.6)
```

where

I = Index (0 through 224)

A = Derived coefficient.

4.1.7 Program SHARMS

Program SHARMS, not shown in the data flow configuration, was designed to predict the parameters of mean sky cover, mean correlation, and population for any given location on the globe using the coefficients derived from program WESPAN. Predicted mean sky covers and mean correlations were in units of equivalent normal deviates which were converted to values between 0.0 and 1.0 by routine PNORM (Appendix A). The program predicted a selected parameter at every five degrees of latitude and longitude through use of equation (11). Each predicted value was stored in an array of 72 by 37 points. Bilinear interpolation between the points was then used to create patterns of predicted parameters that were mapped onto a cylindrical equal area projection of the globe from 90 deg South to 90 deg North. The equal area projection was simply generated by using a linear scale in longitude. Latitude was scaled as sine of the latitude. Resulting maps were used for quality control of the data processing effort throughout the entire project.

Figure 4 depicts the global patterns of percent mean sky cover as predicted from the sky cover coefficients derived from DOE/Burger data for winter at 00 MAST. Figure 5 shows the corresponding mean correlation patterns as predicted from coefficients of mean correlation for the same time period.

Figure 6 shows patterns of predicted percentages of population values for the winter case at 00 MAST. The approximate magnitude of the population for a given pattern can be computed by

$$N = (P+5\%)/100 \times PMAX \tag{17}$$

where

P = Percent of pattern category in Figure 6

PMAX = 990 (Upper limit of sample size chosen for winter cases)

N = Approximate sample size. (Days)

4.2 PROCESSING INTERANNUAL VARIANCE OF MEAN SKY COVER

Interannual standard deviations of mean sky covers were computed from 11 years of DOE monthly mean sky covers. Coefficients of the standard deviations were generated in order to subsequently compute and display the amount of expected error associated with predicted sky cover amounts and other related statistics within the C Cloud S program.

4.2.1 Program FETC2

As shown in Figure 7, a return to the DOE cloud data tape M5550 was necessary in order to extract monthly mean sky covers for computing interannual variances. Beginning with map group number 100, program FETC2 invoked by XEQFE2, extracted from file number 2 the monthly mean sky cover over land areas. A second run was made to extract the same parameter for the grid points over the ocean areas. These were found in file 6 starting with map group number 1340. The extracted data were labeled on disk files with names such as JAL102 (JAnuary Land data from group 102).

Figure 4. Pattern Map of Predicted Percent Mean Sky Cover for the Winter Season at 00 MAST. Predictors Were Derived from the DOE/NCAR and Burger Sky Cover Data Ensembles.

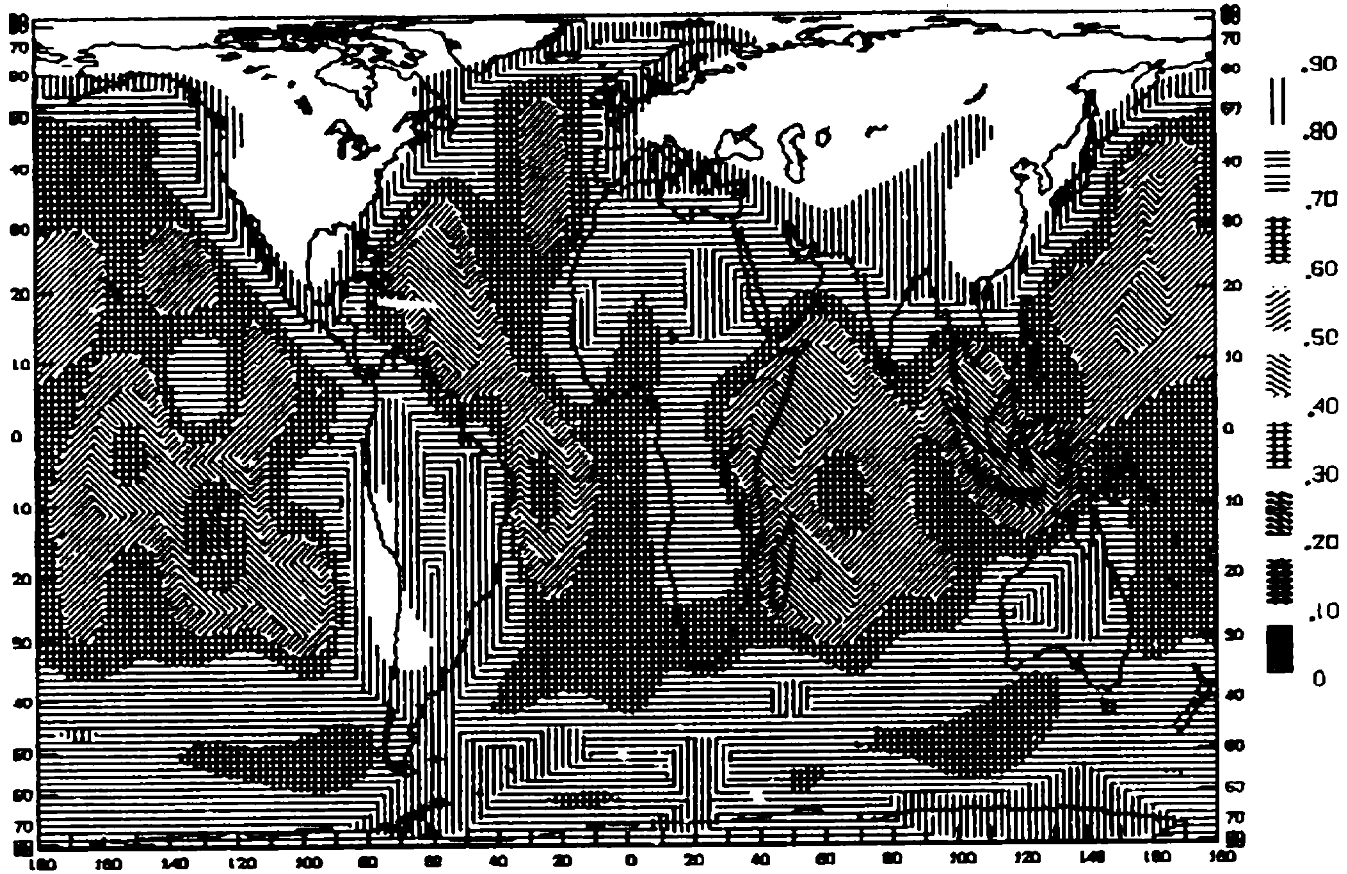


Figure 5. Pattern Map of Predicted Mean Correlation for the Winter Season at 00 MAST. Predictors Were Derived from the DOE/NCAR and Burger Sky Cover Data Ensembles.

Figure 6. Pattern Map of Predicted Population Percentages for the Winter Season at 00 MAST. Predictors Were Derived from the DOE/NCAR and Burger Sky Cover Data Ensembles.

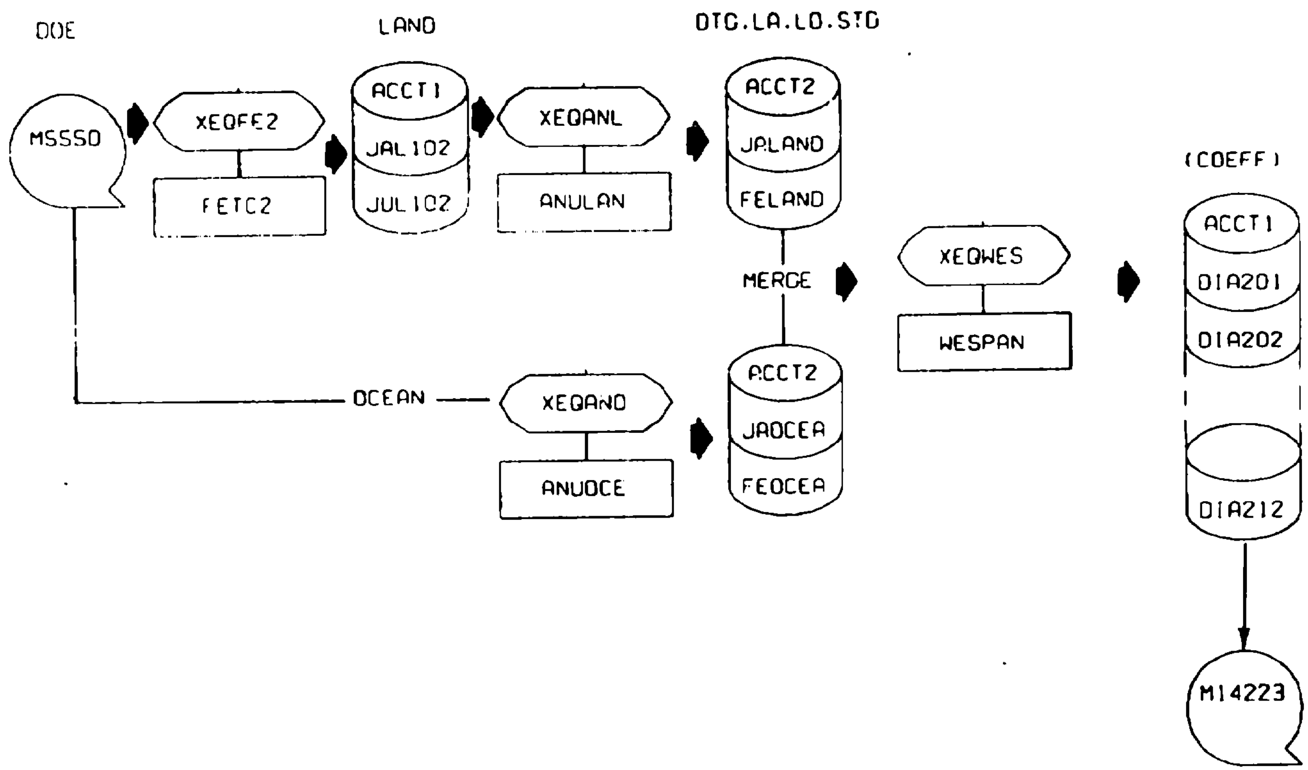


Figure 7. Data Flow Configuration Used in Processing the Interannual Standard Deviations of Mean Sky Cover Observations from DOE Data.

4.2.2 Program ANULAN, AUNOCE

Program ANULAN read the monthly mean sky cover values extracted from FETC2 and summed them for each grid box. The sum of the squares of the sky cover amounts were also computed. At the completion of the process the standard deviation of mean sky cover was computed for each grid box using the basic formula below.

$$s = +\sqrt{\frac{\sum m^2}{n} - \left(\frac{\sum m}{n}\right)^2} \tag{18}$$

where

m = monthly mean sky cover amount

n = number of observations

s = positive standard deviation.

A date time group DTG, latitude-longitude reference, and positive standard deviation values were stored as files JALAND or FELAND, etc., meaning JAnuary or FEbruary LAND standard deviation values. Program ANUOCE handled the ocean data in the same manner. Names like JAOCEA meaning JAnuary OCEAN values were assigned to these data files. At the completion of this process the land data were merged with the ocean data and the land data files purged.

4.2.3 Program WESPAN

Finally, program WESPAN read the interannual standard deviations together with locators and compressed the data into coefficients which were stored on files called DIA201, DIA202, etc., meaning DOE InterAnnual standard deviations, version 2 for January, 01, or February, 02, etc. Twelve sets of coefficients were generated, one set for each month. Tape M14223 was utilized to retain the coefficients and the software.

Figure 8 shows predicted patterns of interannual standard deviations of mean sky covers derived from 11 DOE January monthly sky cover means.

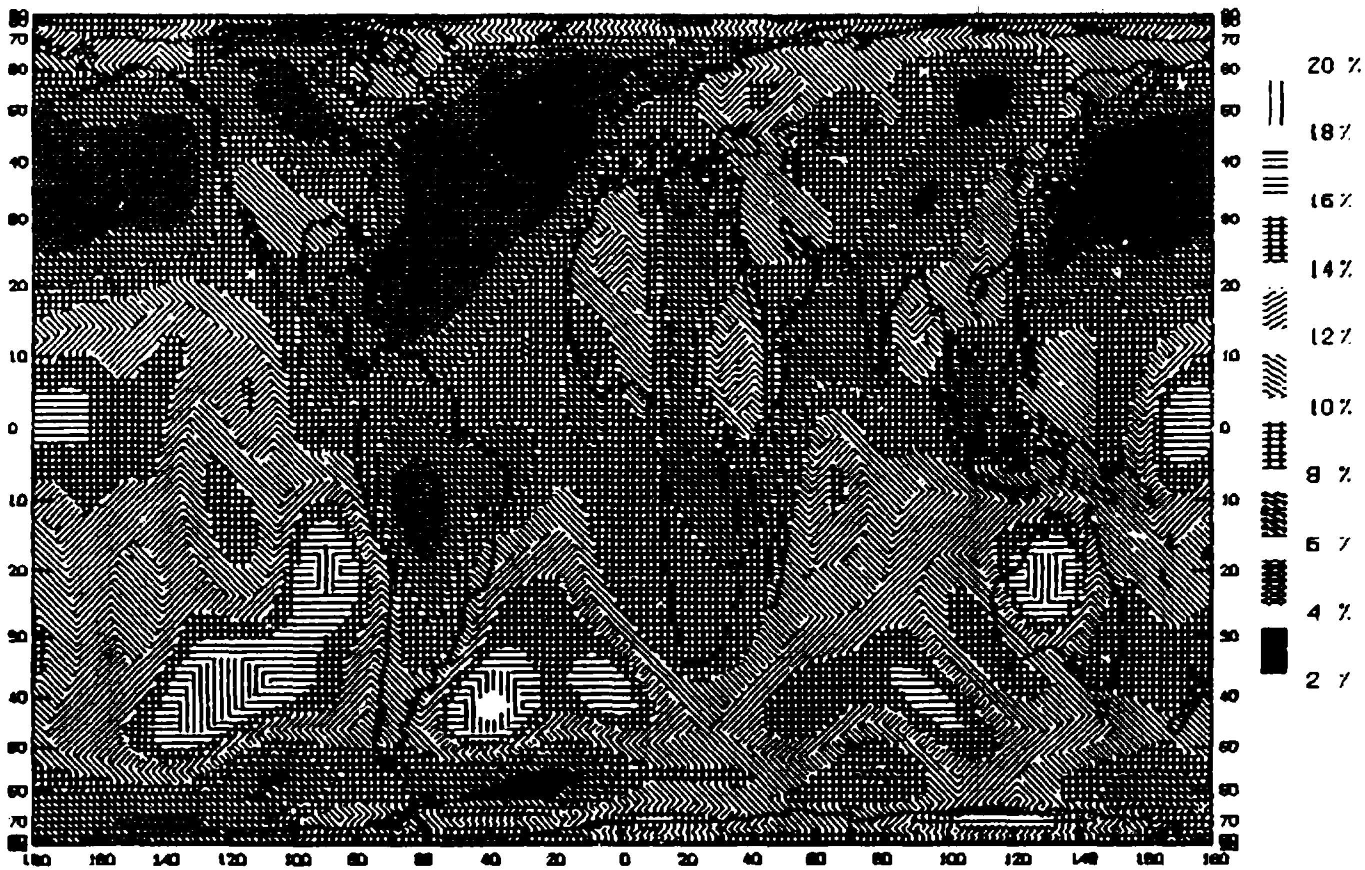


Figure 8. Pattern Map Showing Predicted Percent Interannual Standard Deviations of Mean Sky Covers. Predicted Coefficients Were Derived from DOE January Monthly Mean Sky Cover Data.

5. NIMBUS-7 CLOUD DATA

The NASA NIMBUS-7 cloud database was generated from the Temperature Humidity Infrared Radiometer (THIR) "11.5-micron" radiances together with the Total Ozone Mapping Spectrometer (TOMS)-derived "ultraviolet" reflectivities, climatological temperature lapse rates and concurrent surface temperature and snow-ice information from the Air Force three-dimensional nephanalysis archive. (Hwang, et al. 1988b)

According to Hwang, et al. (1988a), the primary merits of the NIMBUS-7 Cmatrix Cloud dataset are 1) observed radiances used for this cloud dataset were obtained from the same instruments used over six continuous years between 1979–1985, so it represents the most homogeneous satellite derived cloud dataset available; 2) the cloud dataset is efficiently retrieved and stored, so that the analysis of the global cloud distribution over climatological time scales is feasible; and, 3) daily as well as monthly averages and variances of cloud data are presented together with a correlation coefficient surface temperature archive. Because of these merits, this cloud database was acquired to further augment cloud statistics within the development of the C Cloud S program.

The NIMBUS-7 Cmatrix cloud data were received on seven magnetic tapes. The data on these tapes consisted of more than 100 parameters that were archived twice a day at noon and midnight. Among parameters archived are daily means of the percentages of cloud cover, the percentage of total low, middle, and high cloud layer amounts, and clear air radiances, and their spatial and temporal variances. These parameters, identified by number, are computed for 2070 NIMBUS-7 earth radiation budget (ERB) target areas. Each target represents an area of approximately 500 x 500 kms. A typical spatial distribution of data within the 2070 ERB target areas can be seen in Figure 9. Every other south-to-north satellite pass is purposely left blank to enhance the actual data positioning within each swath. The numbers in the figure represent Greenwich Mean Time (GMT) for the data within each ERB target for an orbital ascending node at noon time.

5.1 PROCESSING NIMBUS-7 DATA

The major portion of the data extracted from this database was composed of daily mean cloud covers, variances, and populations within each ERB target for each of twelve months. Figure 10 illustrates the flow of events used in the processing of the NIMBUS-7 Cmatrix Cloud data.

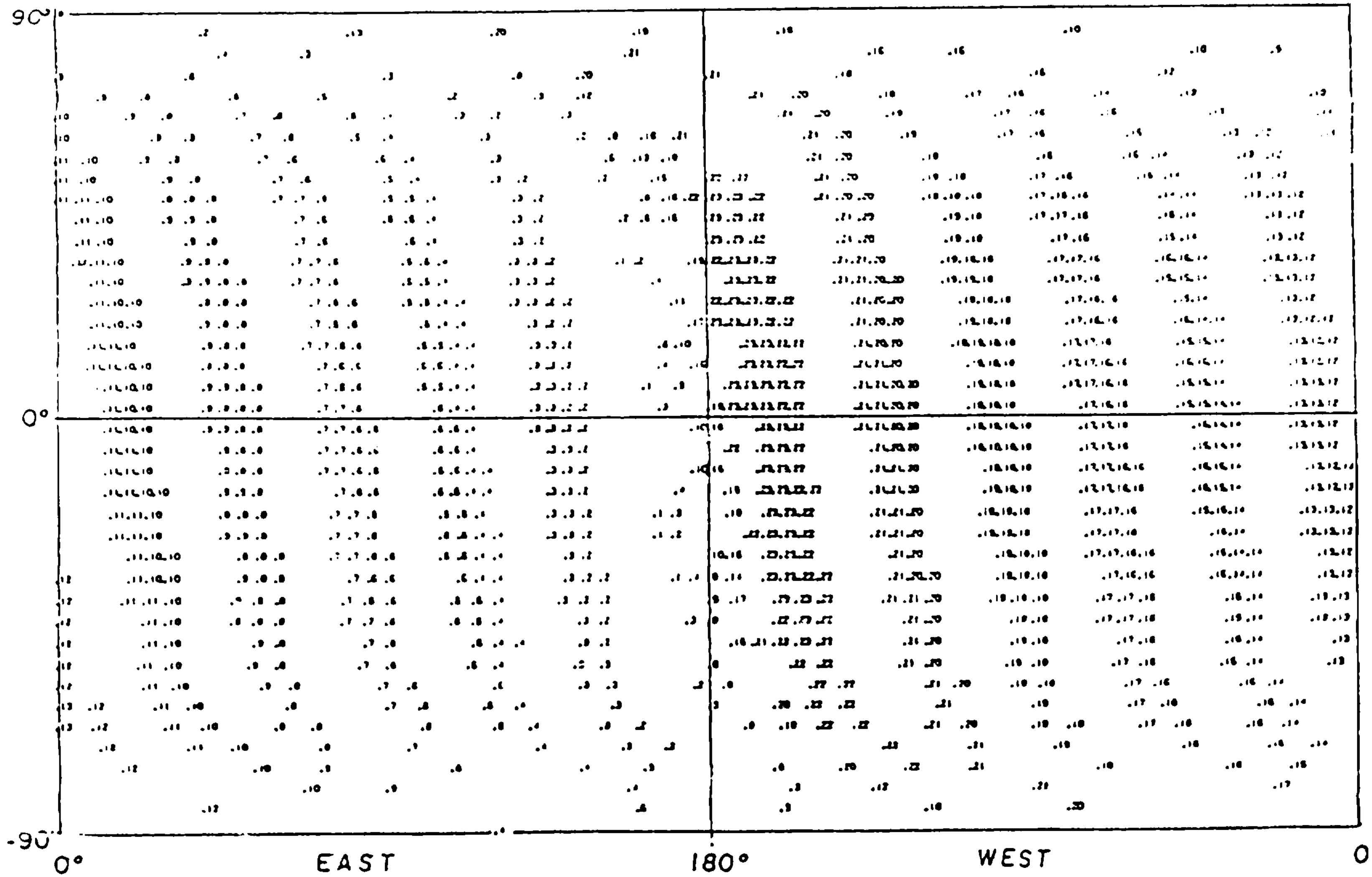


Figure 9. A Track Map of NIMBUS-7 CMATRIX Data Dissemination Over 2,070 ERB Target Areas. Every other Orbital Swath is Purposely Left Blank to Show the Orientation of South to North Noon Time Satellite Coverage. Numbers in the Figure Represent GMT's of Data Observations.

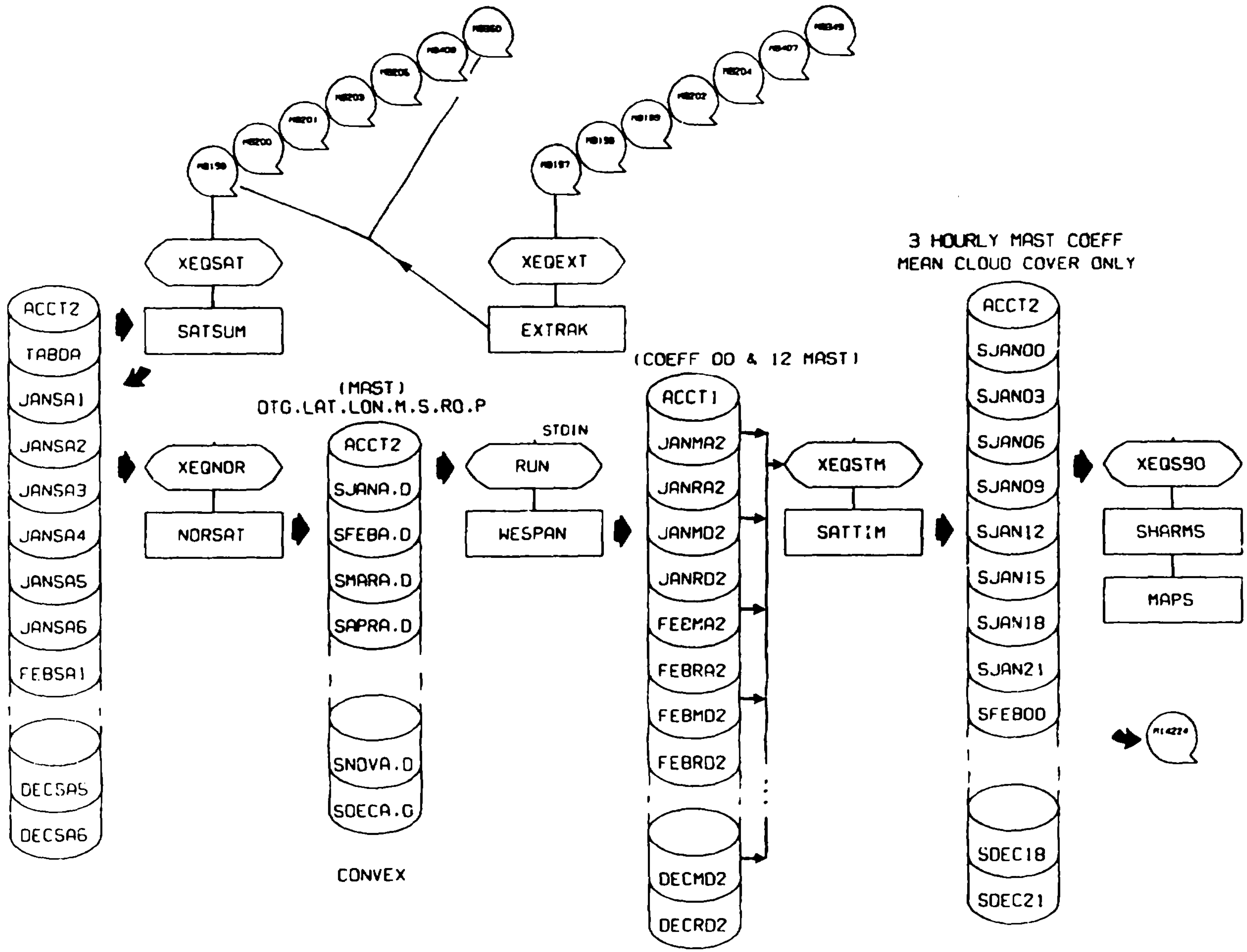


Figure 10. Data Flow Configuration Used in Processing the NIMBUS-7 Cloud Cover Climatology.

5.1.1 Program EXTRAK

Procedure XEQFET calls on program EXTRAK to extract from a given Cmatrix data tape the daily maps of parameters selected by the parameter number defined in the Cmatrix tape user documentation by Wellemeyer (1986). A list of the tape numbers assigned to the acquired Cmatrix data tapes and their archived period of record are listed in Table 3. Table 4 lists the parameter definitions of the parameter number codes selected for data extraction. The extracted daily parameter maps for all months and from each Cmatrix tape are stored on a second set of tapes for further processing.

Table 3. CMatrix Tape Numbers and Archived Period of Record

Tape Number	Period of Record
M8197	Apr 79–Oct 79
M8196	Nov 79–Oct 80
M8199	Nov 80–Oct 81
M8202	Nov 81–Oct 82
M8204	Nov 82–Oct 83
M8407	Nov 83–Oct 84
M8349	Nov 84–Mar 85

Table 4. Parameter Definition Given Selected CMatrix Parameter Number Code

Number Code	Definition
1	Recommended total percent cloudiness ascending
2	RMS of recommended total percent cloudiness ascending
23	Mean subtarget area population ascending
25	GMT (minutes) ascending
55	Total percent cloudiness descending
56	RMS of total percent cloudiness descending
75	Mean suntarget area population descending
77	GMT (minutes) descending

5.1.2 Program SATSUM

Program SATSUM is utilized to sum the daily map data stored on the new tapes for a selected month and parameter code. The program uses a navigation table called TABDA, published by Wellemeyer (1986), to navigate the 2070 NIMBUS-7 ERB target areas. Daily sums of 4 parameters over individual months for a period of record of six years were stored in binary form on files named JANSA1 and JANSA2, ..

JANSA6 meaning January Sums for years 1,2, ... 6. The four parameters summed are 1) mean cloud cover, 2) mean cloud cover squared, 3) GMT's and 4) number of observations.

5.1.3 Program NORSAT

Program NORSAT reads in the binary data summation files from SATSUM and computes the mean and standard deviation of cloud cover in each ERB target area for both ascending and descending nodes using equation 18 where

m = sum of satellite mean cloud cover

n = satellite population count

s = positive standard deviation

Given the above information, mean correlation is also computed using the methods in 4.1.2. A date time group, latitude-longitude locator, mean cloud cover, standard deviation, mean correlation, and population for each ERB target area are then written to permanent files using the standard format 3.1.1. Names of the files written were SJANA, SFEBA, ... SDECA for monthly ascending node data and SJAND ... for descending node data.

5.1.4 Program WESPAN

The files created by program NORSAT above are transported to the UNIX system where program WESPAN is utilized for compressing the satellite data into coefficients. The coefficients generated are for mean cloud cover and mean correlation at 00 and 12 MAST. Coefficients are not generated for the satellite population values since the sample sizes within all ERB targets are reasonably constant at about 186 or 31 days x 6 years.

Files containing the satellite coefficients are named JANMA2 and JANRA2 meaning January mean cloud cover (M) and January mean correlation (R) for ascending (A) node, version 2, and JANMD2 or JANRD2 for the descending node. The January file is followed by files for FEB, MAR, etc.

Near the completion of each WESPAN computer run, a multiple correlation coefficient is computed and printed to examine how well the generated coefficients fit the data. The multiple correlations are tabulated for each run. They are presented in Table 5 for NIMBUS-7, DOE, and DOE interannual standard deviations.

5.1.5 Program SATTIM

The NIMBUS-7 satellite cloud database offers excellent spatial coverage of cloud amount statistics over the globe. It gives good estimates of cloud coverage over areas where surface observations are sparse or completely void. However, the data are collected only twice a day, noon and midnight. Therefore, we developed program SATTIM to interpolate the spatial coefficients derived from the satellite data for predicting cloud amounts at 00 and 12 MAST to eight times of day (00, 03, 06, ... 21 MAST). To accomplish the interpolation for each coefficient A, we use a first order Fourier, equation 19, with a temporal phase adjustment of 2 hours, and an amplitude adjustment, of the amount computed in equation 20 below.

$$A_i(t) = \frac{(a_i + b_i) + h(a_i - b_i)\cos\left[\frac{2\pi(t-2)}{24}\right]}{2} \tag{19}$$

where

$$h = \frac{1}{\cos(30^\circ)} \tag{20}$$

t = MAST(00,03,06,... or 21)

a_i = 00 MAST cloud cover coefficients

b_i = 12 MAST cloud cover coefficients

A_i = coefficient for mean cloud cover

i = Index 0 through 224.

Table 5. Multiple Correlation Coefficients[1]

Month NIMBUS	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
$\bar{M}$ (A)	.945	.951	.952	.951	.952	.955	.953	.956	.961	.957	.948	.945
(D)	.947	.951	.948	.940	.937	.954	.955	.955	.955	.954	.942	.943
$\bar{R}$ (A)	.851	.858	.870	.866	.866	.870	.889	.884	.883	.874	.850	.835
(D)	.869	.881	.887	.877	.886	.870	.847	.832	.852	.867	.835	.841
DOE												
00	.901			.890			.919			.921		
03	.900			.913			.931			.924		
06	.896			.914			.920			.921		
$\bar{M}$ 09	.913			.928			.939			.935		
12	.900			.906			.932			.926		
15	.913			.920			.943			.931		
18	.898			.892			.931			.913		
21	.905			.918			.940			.928		
00	.872			.858			.832			.872		
03	.903			.905			.880			.900		
06	.875			.887			.847			.893		
$\bar{R}$ 09	.921			.926			.895			.918		
12	.909			.900			.832			.895		
15	.923			.911			.854			.907		
18	.892			.865			.807			.868		
21	.916			.901			.873			.904		
00	.778			.783			.766			.771		
03	.799			.811			.803			.797		
06	.779			.778			.772			.781		
P 09	.817			.816			.810			.809		
12	.788			.790			.776			.779		
15	.806			.808			.804			.800		
18	.788			.796			.781			.785		
21	.806			.820			.802			.800		
DOE												
IA	.770	.775	.766	.724	.727	.744	.728	.720	.741	.765	.745	.770

[1]Listed are multiple correlation coefficients computed by program WESPAN for monthly NIMBUS 7 mean (M) cloud covers and mean correlations (R) derived from data over ascending nodes (A) and descending nodes (D). Multiple correlation coefficients are also listed under mid season months that were derived for the eight GMT seasonal DOE mean sky covers (M), mean correlations (R), and populations (P). Finally, multiple correlation coefficients derived from DOE monthly interannual standard deviations are listed in the row labeled IA.

Each interpolated set of 225 coefficients for predicting mean cloud coverage at 3 hourly time intervals is stored on disk files with names like SJAN00 or SFEB12, meaning Satellite JAnuary 00 or FEbruary 12 MAST. The resulting new ensemble of coefficients was then in a form suitable for the blending of satellite mean cloud cover predictors with ground based mean sky cover predictors from Burger and DOE data. (see Section 7).

5.1.6 Program SHARMS

Program SHARMS used the coefficients generated by WESPAN to produce global pattern maps of predicted satellite observed mean cloud covers and mean correlations for data quality control. An example of the predicted global mean cloud cover for January at 00 MAST is displayed in Figure 11. Mean correlation for January at 00 MAST is shown in Figure 12. Tape M14224 was used to back up all software and files generated during the satellite data processing effort.

6. DATSAV DATA

This dataset contains up to 11 years of hourly surface observations taken at individual stations around the world. The data have been formatted and archived on magnetic tape by the U.S. Air Force Environmental Technical Applications Center (USAFETAC). This archive of surface observations is called DATSAV. The archived hourly data consists of weather conditions such as temperature, winds, dew points, total sky covers, cloud types, etc. Documentation of the DATSAV database can be found in the DATSAV Reference Manual, 1977. A more recent updated document is the DATSAV2 Reference Manual, 1986.

6.1 PROCESSING DATSAV DATA

DATSAV data tapes for 66 stations positioned in clusters at various locations around the world were successfully processed by Willand (1988) in order to archive hourly mean sky cover and scale distance parameters. Since then 26 additional stations within new clusters were added to augment this data collection. These new samples were in the updated DATSAV2 format. Figure 13 depicts the general positions of all the acquired DATSAV clusters. Table 6 lists names and locations of the stations within each cluster. The data flow for processing the DATSAV data is shown in Figure 14.

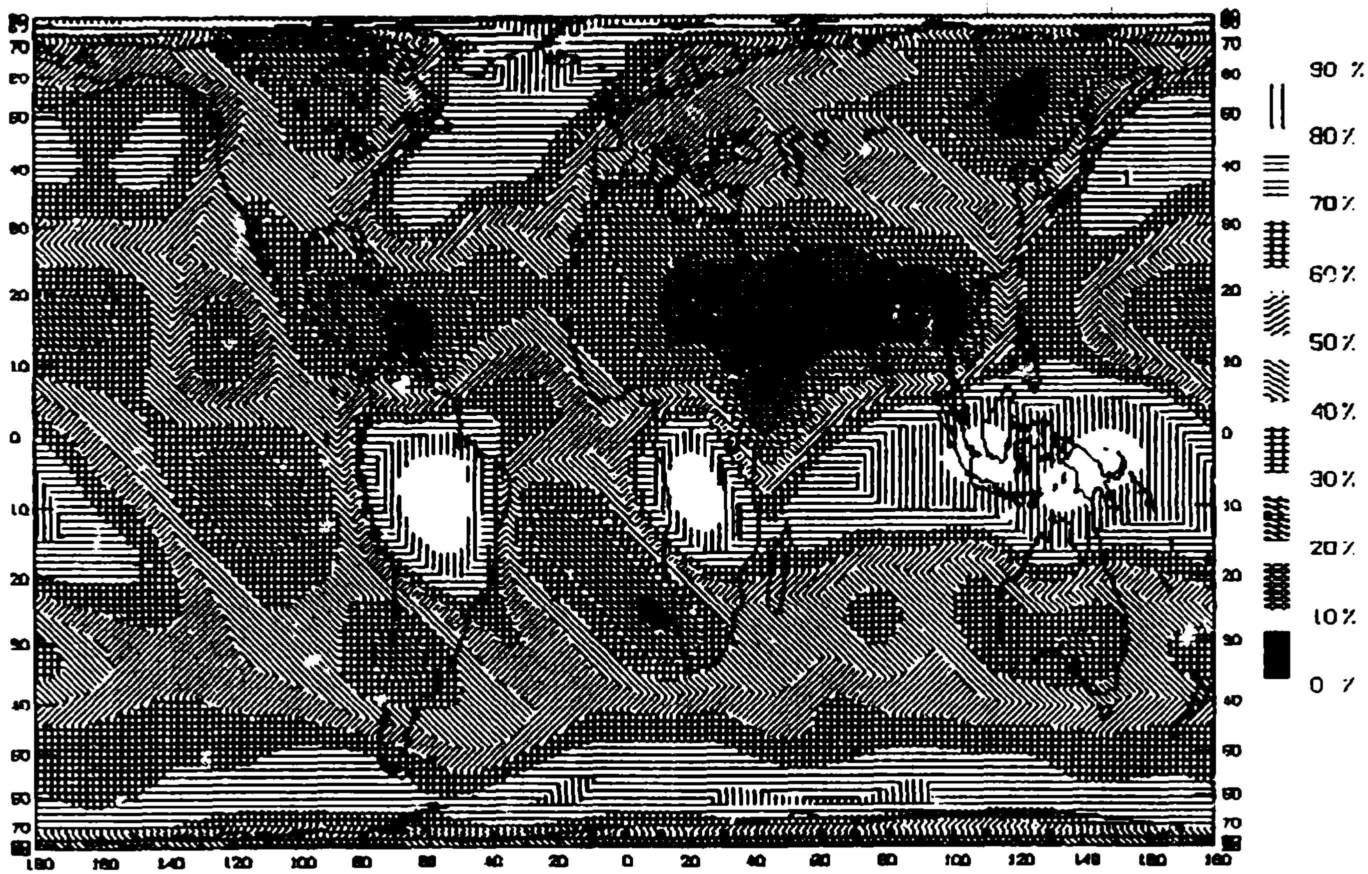


Figure 11. Pattern Map of Predicted Percent Mean Cloud Cover for January at 00 MAST. Predictors Were Derived from the NIMBUS-7 CMATRIX Cloud Cover Climatology.

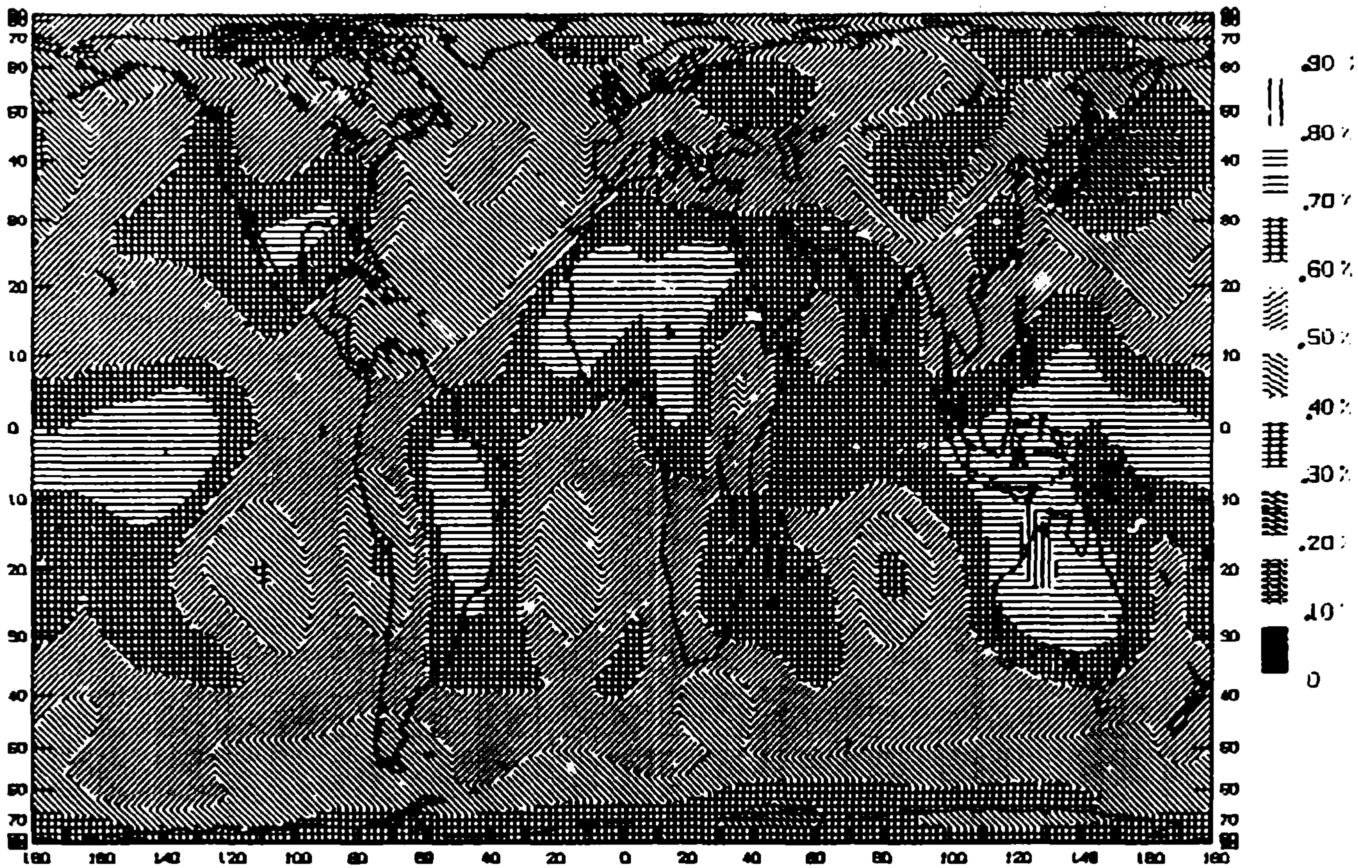


Figure 12. Pattern Map of Predicted Mean Correlation for January at 00 MAST. Predictors Were Derived from the NIMBUS-7 CMATRIX Cloud Cover Climatology.

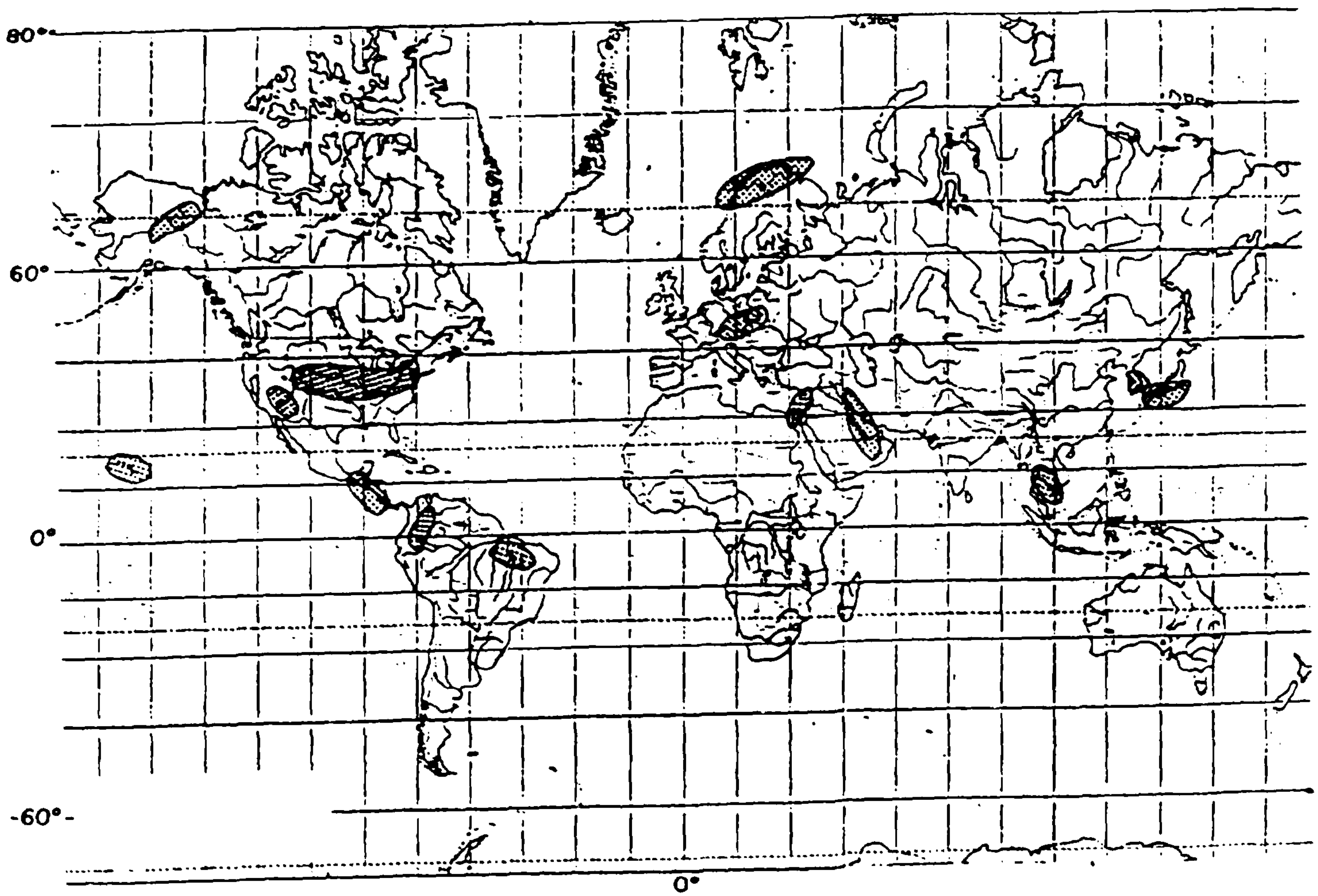


Figure 13. Map of Cluster Positions for DATSAV and DATSAV2 Data Collections.

Table 6. Listing of DATSAV Stations Within Clusters

Sta.	Lat.	Lon.	Elev.	Name
MIDWEST:				
72638	44.37	-84.68	0351	Houghton Lake, MI
72652	43.07	-95.53	0423	Pickstowne, SD
72424	37.90	-85.97	0233	Ft. Knox, KY
72551	40.85	-96.75	0362	Lincoln, NE
72640	42.95	-87.90	0220	Milwaukee
72641	43.13	-89.33	0262	Madison
72637	42.97	-89 73	0238	Flint
72552	40.97	-98 32	0563	Grand Isle, NE
72556	41 98	-97.43	0479	Norfolk, NE
72423	38.18	-85.73	0151	Louisville, KY
72635	42.88	-85.52	0242	Grand Rapids, MI
72636	43.17	-86.23	0191	Muskegon, MI
72529	43.12	-77.67	0171	Rochester, NY
72528	42.93	-78.73	0217	Buffalo, NY
72545	41.88	-91.70	0263	Cedar Rapids, IA
72546	41 53	-93.67	0292	Des Moines, IA
72734	46.47	-84.37	0220	Salt St. Marie
SOUTHWEST:				
725690	42.55	-106.128		Casper, WY
746120	35.41	-117.41	0681	China Lake, CA
723865	36.25	-115.03	0570	Nellis AFB
724660	38.82	-104.72	1881	Colorado Springs, CO
722800	32.40	-114.36	0063	Yuma, AZ
726940	44.92	-123.02	0061	Salem/McNary
MIDDLE EAST I:				
400800	33.31	36 81		Damascus, Syria
401840	31.46	35.14		Jerusalem, Israel
623660	30.00	31.17		Cairo, Egypt
MIDDLE EAST II:				
404160	26.18	50.08		Dharan, S. Arabia
404380	24.31	46 47		Ridyadh, S. Arabia

Table 6. Listing of DATSAV Stations Within Clusters (Continued)

Sta.	Lat.	Lon.	Elev.	Name
MIDDLE EAST II:				
405820	29.04	47.59		Kuwait, Kuwait
406890	30.30	47.47		Basrah, Iraq
ASIA I:				
471220	37.11	127.04	0041	Osan A.B., S. Korea
470580	39.03	125.48		Pyong Kang, N. Korea
ASIA II:				
476420	35.45	139.21		Yokota, Japan
478270	31.24	130.39		Kagoshima, Japan
ASIA III:				
484560	13.50	100.29	0004	Bangkok, Thailand
488200	21.04	105.50		Hanoi, Vietnam
488550	16.08	108.22		Da Nang, Vietnam
489000	10.46	106.34		Ho Chi Minh, Vietnam
983270	15.11	120.32		Clark AFB, Phillipines
466960				Tapei, Taiwan
CENTRAL AMERICA:				
786410	14.37	-90.32		Guatemala Cita, Guatemala
786630	13.45	-89.11		San Salvador, El Salvador
787200	14.08	-87.15		Tegucigalpa, Honduras
787410	12.10	-86.16		Managua, Nicaragua
787600	10.00	-85.00		Punta Arenas, Costa Rica
783060	8.92	-79.60		Howard AFB, Panama
SOUTH AMERICA:				
90370	0.87	-77.68	2975	Ipiales
80110	6.22	-75.60	1498	Medellin
80089	7.00	-74.72	0610	Amalfi Antioquis
80259	3.55	-76.38	0969	Cali/Parameseca
80308	2.45	-76.60	1730	Popa Van/Leon-Valei

Table 6. Listing of DATSAV Stations Within Clusters (Continued)

Sta.	Lat.	Lon.	Elev.	Name
AMAZON BASIN:				
82678	-6.17	-42.58		Floriano/Cangapara
82562	-5.21	-49.07		Maraba
HAWAII:				
91186	21.15	-157.10	0138	Molokai
91182	21.35	-157.93	0005	Honolulu
SWEDEN-NORWAY:				
01006	69.07	15.18	0014	Anda
01025	69.68	18.92	0010	Troniso
01059	70.07	24.98	0008	Banak
01041	69.75	21.03	0004	Nororeisa
10570	69.37	24.43	0285	Cuoutatmauki
10490	69.98	23.35	0003	Lufthaun
01098	70.37	31.10	0014	Vardo
20960	67.20	23.42	0168	Fajala
22101	68.90	28.45	0118	Virtaniesaii
28360	67.37	26.65	0178	Sodankyla
22106	68.60	31.80	0034	Padun
22119	68.32	33.33	0130	Pulozem
22204	67.55	30.43	0221	Kordor
22205	67.60	31.17	0160	Ena
22216	67.55	33.39	0136	Apatity
22232	67.70	37.12	0244	Kolmyarr
22249	67.13	39.67	0149	Kaneuka
GERMANY:				
10147	53.63	9.98	0016	Hamburg
67500	46.65	8.62	2287	Guetsch
67620	46.17	8.88	0198	Lacarno
67700	46.00	8.97	0276	Logiano
10235	53.00	9.83	0077	Soltau
10803	48.02	7.85	0300	Freiburg
10908	47.87	8.00	1493	Feldberg

Table 6. Listing of DATSAV Stations Within Clusters (Continued)

Sta.	Lat.	Lon.	Elev.[1]	Name
GERMANY:				
16066	45.62	8.73	0211	Miland
16072	45.87	9.07	1319	Monte Bisino
16080	45.43	9.28	0107	Milano Linate
06700	46 25	6.13	0436	Geneu/Cointrin
06670	47.48	8.53	0432	Zuritch
06680	47.25	9.35	2500	Saentis
06720	46.22	7.33	0482	Sion
06760	46.17	8.78	0379	Locamo
06610	46.82	6 95	0489	Payern
06990	47.13	9.53	0463	Vaduz
ALASKA:				
70174	66.92	-151.52	0205	Bettles
70063	70.15	-148.20		Prudhoe
70265	64.40	-147.7	0128	Eielson

[1]Elev. = Meters

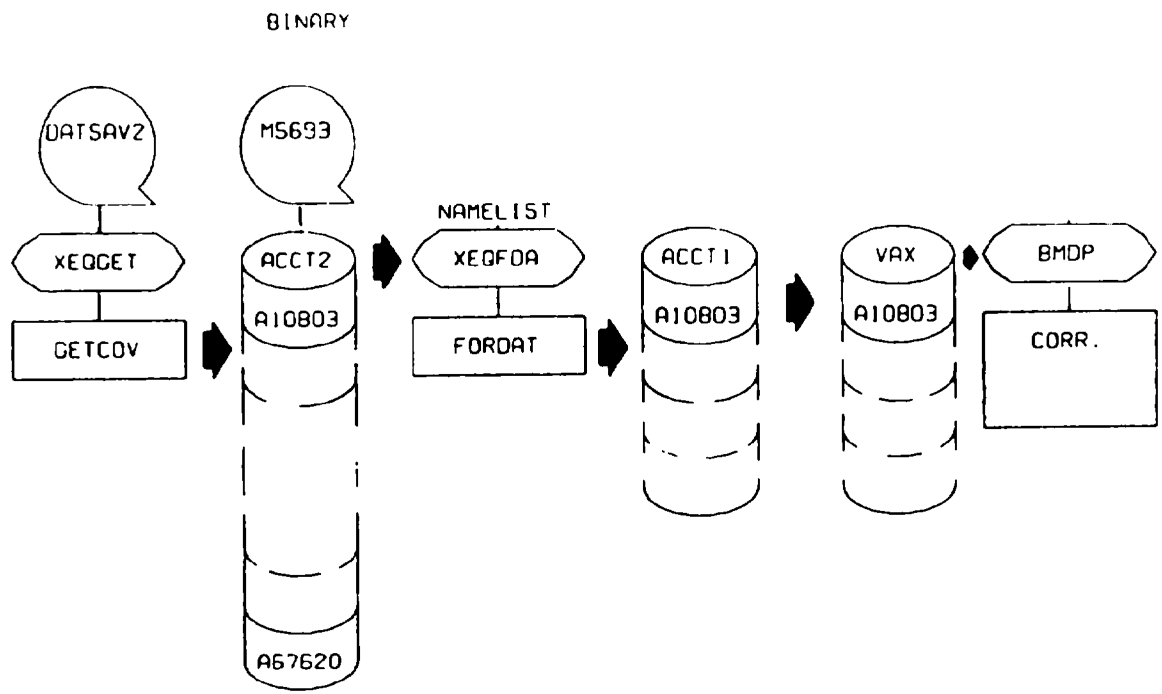


Figure 14. Data Flow Configuration for Formatting DATSAV Data for Subsequent Processing Using the BMDP Program.

6.1.1 Program GETCOV

The flow of data processing begins with the newly acquired DATSAV2 data. Procedure XEQGET positions the DATSAV2 data tape to the proper file so that program GETCOV can extract the hourly sky cover observations from a selected station. These hourly sky cover observations from each selected station are packed and stored on disk as binary random access data files. Detailed packing and storage documentation of the sky cover observations can be found in Willand (1988). After processing the newly acquired DATSAV2 data through program GETCOV, backup tape M5693 containing DATSAV hourly sky cover observations (processing described by Willand (1988)) are spooled back onto disk storage. Hourly sky cover data for all the selected stations in each cluster are then ready for further processing.

6.1.2 Program FORDAT

Program FORDAT was written to format DATSAV sky cover data into an acceptable form for processing using the Biomedical Data Processing Statistical Software Package (BMDP) (Dixon, et al., 1988). The BMDP has an accurate tetrachoric correlation package which is useful for defining spatial and temporal correlations of sky cover conditions over stations within the DATSAV clusters.

Procedure XEQFDA, used for executing program FORDAT, requires a FORTRAN NAMELIST input file to define the period of record selected for formatting. The necessary NAMELIST arguments are:

```
$IDEF
     PRINT = .TRUE.,  Print resulting data values
     IY1 = 73,        First year of data extraction
     IY2 = 83,        Second year of data extraction
     IM1 = 7,         First month of data extraction
     IM2 = 7,         Last month of data extraction
   NUMT = 12,         Pointer to output file
$END
```

The format used by program FORDAT to write the new data files for BMDP processing is as follows.

```
WRITE(NTAPE,100)ISTA,XLAT,XLONG,IY,IM,ID,(IDATA(I,ID),I=1,IT1,IT2)
100 FORMAT(I6,F6.2,F7.2,3I2,24I2)
```

where

ISTA = Station number

XLAT = Station latitude

XLONG = Station longitude

IY = Year

IM = Month

ID = Day

IT1 = Start time

IT2 = End time

The formatted ASCII data files are first stored on the CYBER system and then transmitted to the VAX system which processes spatial and temporal correlations of sky cover using the BMDP package. The resulting spatial and temporal correlations are then represented graphically to define the decay of conditionality over space and time within each cluster. Figure 15 shows the temporal decay of sky cover derived from 16 stations over the mid-western United States for January.

7. DATABASE BLENDING

At this point of the data processing activities, the FINARC portion of Burger data is merged with the DOE database and the determination of the coefficients is completed. The coefficients predict mean sky covers, mean correlations, and populations for up-looking sky cover conditions at three hourly intervals in MAST for the four seasons. Also completed are the coefficients for predicting mean cloud covers and mean correlations from down-looking satellite observations for the same three hourly intervals but for each of 12 months. The next step in the processing converts DOE seasonal sets of coefficients to monthly sets and in the process blends together coefficients for predicting ground observed mean monthly sky covers with those for predicting satellite observed mean monthly cloud covers. The flow of the blending process is shown in Figure 16 and details of the processing are discussed below.

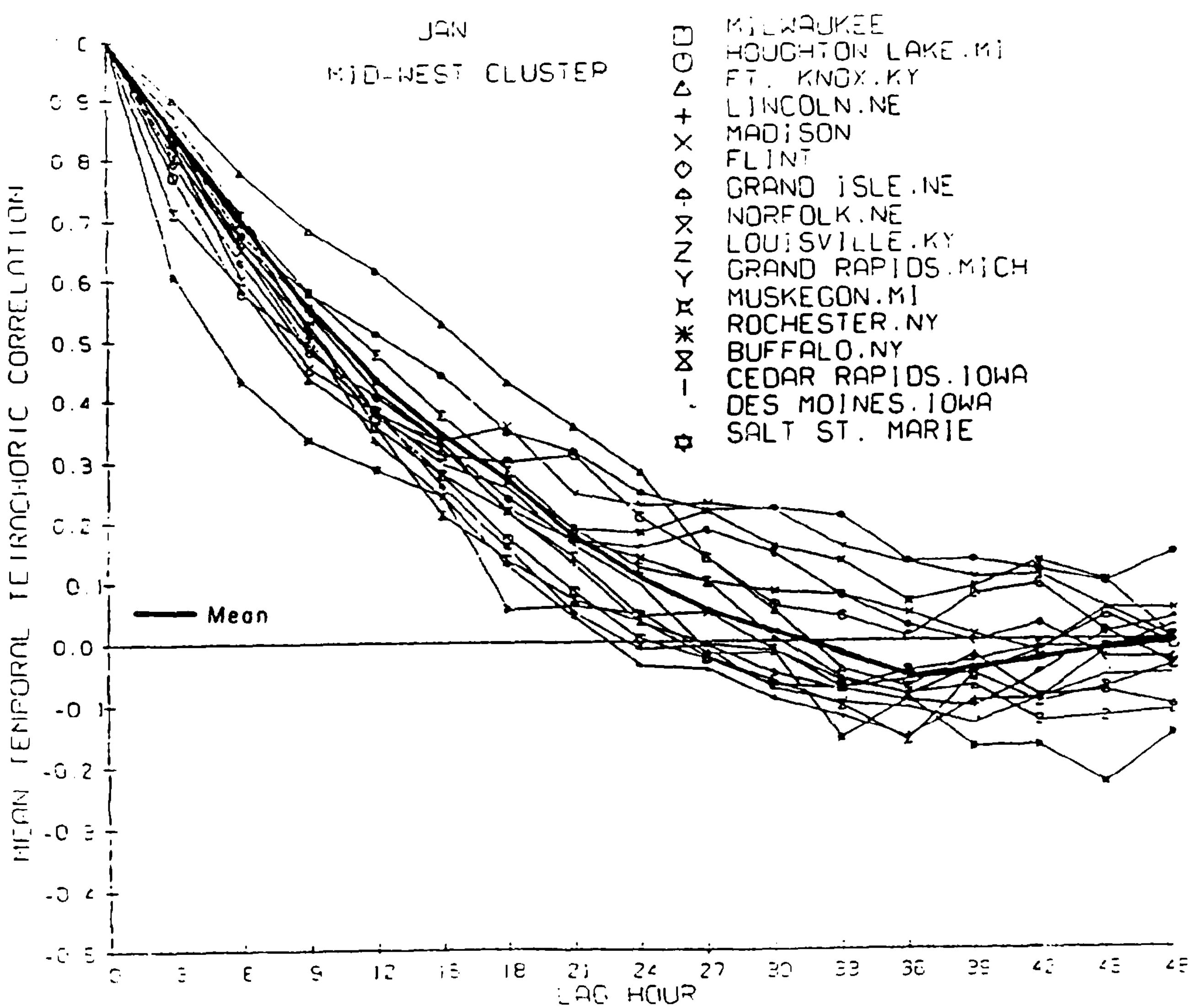


Figure 15. Temporal Correlations of Sky Cover for the Midwestern U.S. Cluster.

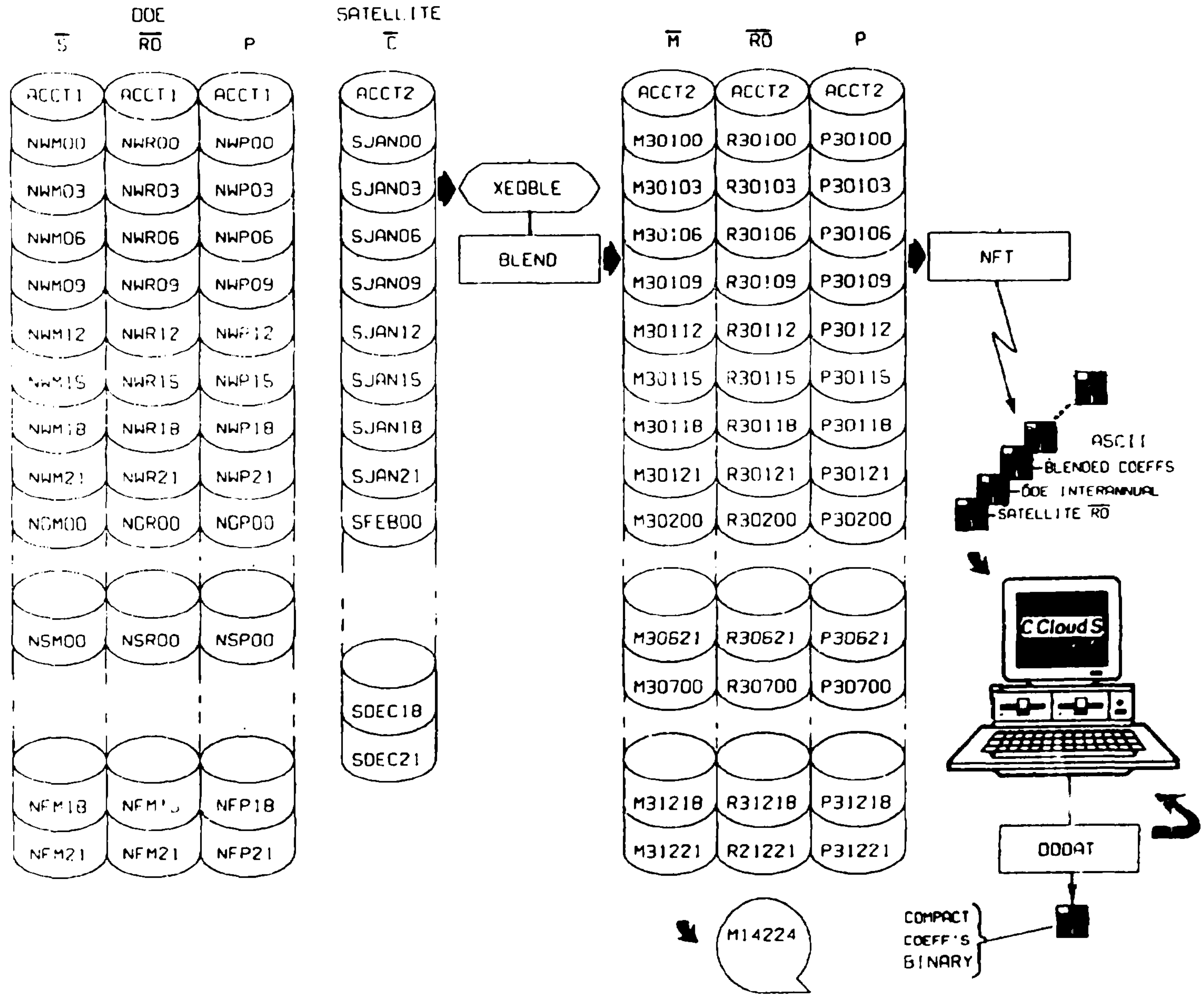


Figure 16. Data Flow Configuration Used to Blend Coefficients Derived from DOE Data With Those from NIMBUS-7 Data.

7.1 PROGRAM BLEND

Initially, program BLEND reads in all sets of coefficients needed to achieve the blending process. This involves reading and storing into the program the coefficients for DOE mean sky covers (S), mean correlations ($\overline{\mathrm{RO}}$) and populations (P) for all four seasons at all eight 3 hourly MAST's. Also read and stored are all the coefficients for satellite mean cloud covers ($\overline{\mathrm{C}}$) adjusted for all 8 times of day for each of the 12 months, (Section 5.1.5, equation 19). The following computations are then performed on all the incoming coefficients.

First, coefficients derived from DOE and Burger climatologies for predicting mean sky cover, mean correlation, and population by <u>seasons</u> are converted to coefficients capable of predicting the three parameters by <u>months</u>. This is accomplished through use of equations 21 and 22.

$$x = \frac{2\pi(M-1)}{12} \tag{21}$$

and

$$s_i = \frac{(W_i+G_i+S_i+F_i)+(W_i-G_i+S_i-F_i)\cos(2x)}{4} + \frac{(G_i-F_i)\sin(x)+(W_i-S_i)\cos(x)}{2} \tag{22}$$

where

M = Month (1–12)

W = Coefficient for Winter

G = Coefficient for Spring

S = Coefficient for Summer

F = Coefficient for Fall

s = Coefficient for month M.

i = Index 0 through 224.

Each coefficient for predicting monthly mean <u>sky</u> cover is then blended with the corresponding coefficient for predicting monthly mean <u>cloud</u> cover from satellite climatologies using equation 23.

$$a_i = w_1 s_i + w_2 A_i \qquad (23)$$

where

s_i = Computed monthly DOE coefficient for mean sky cover (equation 22).

A_i = Corresponding satellite coefficient for mean cloud cover (equation 19).

w_1 = .6470588

w_2 = .3529412

a_i = Blended coefficient.

i = Index 0 through 224.

We derived the weights (w_1 and w_2) in equation 23 by setting a maximum sample size for a DOE data sample to be 341 days (31 days x 11 years) and a maximum satellite data sample size to be 186 days (31 days x 6 years). Total sample size is then 527 days (186 + 341). Thus, 341/527 equals .6470588, which is the weight w_1 that is applied to each DOE monthly mean sky cover predictor S_i. The weight w_2 applied to each coefficient to predict satellite monthly mean cloud cover A_i is (1.0–w_1) or .3529412.

Mean correlations derived from satellite observed cloud amounts do not correlate well with those derived from ground observed sky covers. Therefore, coefficients for predicting monthly mean sky dome mean correlations are simply the results of equation 22.

Coefficients computed from equation 22 for predicting population values were blended with the satellite population predictors by simply adding the ratio of 186 days/341 days or .545454 to the first coefficient.

Figure 17 depicts patterns of percent sky cover predicted from the blended coefficients for January at 00 MAST.

Note that coefficients for predicting mean correlation are not blended or weighted but remains as coefficients computed directly from equations 21 and 22.

Final coefficients are written to permanent storage with names such as M30100, meaning coefficients for predicting monthly mean sky covers $\overline{\text{M}}$ version number 3 for January 01 at 00 MAST. $\overline{\text{R}}$ and P are used as prefixes to the names of monthly mean correlation and population coefficients, respectively. Backup tape M14224 is used to store program BLEND and all the blended coefficients.

Figure 17. Pattern Map of Predicted Percent Mean Sky Cover for January at 00 MAST. Predictors Were Derived from Blending the Coefficients for Predicting Sky Cover (DOF/Burger Data) With Those for Predicting Cloud Cover (NIMBUS-7 data).

7.2 PROGRAM NFT

NFT is a network file transfer utility that is utilized to transfer final coefficients from the mainframe to floppy disk storage. The data stored on these floppy's are in ASCII code. Their format is described in format 4.1.6.

Specifically, final version 3 of blended coefficients for predicting mean sky covers M, and the coefficients for mean correlations R, and populations P, are organized on floppy disks by month and time in the following manner.

Disk 1	M30100,03...21	Disk 2	M30300,03...21	Disk 3	M30500,03...21
	R30100,03...21		R30300,03...21		R30500,03...21
	P30100,03...21		P30300,03...21		P30500,03...21
	M30200,03...21		M30400,03...21		M30600,03...21
	R30200,03...21		R30400,03...21		R30600,03...21
	P30200,03...21		P30400,03...21		P30600,03...21
Disk 4	M30700,03...21	Disk 5	M30900,03...21	Disk 6	M31100,03...21
	R30700,03...21		R30900,03...21		R31100,03...21
	P30700,03...21		P30900,03...21		P31100,03...21
	M30800,03...21		M31000,03...21		M31200,03...21
	R30800,03...21		R31000,03...21		R31200,03...21
	P30800,03...21		P31000,03...21		P31200,03...21

Coefficients for predicting DOE (D) interannual (IA) standard deviations (version 2) of sky cover are stored on a single floppy disk number 7 for each month 01 through 12 as:

Disk 7	DIA201	DIA202	DIA203	DIA204	DIA205	DIA206
	DIA207	DIA208	DIA209	DIA210	DIA211	DIA212

Finally, the coefficients for predicting satellite mean correlation are stored on a single floppy disk number 8 in the following order. APRRA2 represents month of April, R is mean correlation, A is ascending node, and D is descending node for version 2.

Disk 8	APRRA2	AUGRA2	DECRA2	FEBRA2	JANRA2	JULRA2
	JUNRA2	MARRA2	MAYRA2	NOVRA2	OCTRA2	SEPRA2
	APRRD2	AUGRD2	DECRD2	FEBRD2	JANRD2	JULRD2
	JUNRD2	MARRD2	MAYRD2	NOVRD2	OCTRD2	SEPRD2

7.3 PROGRAM DODAT

Program DODAT was written in Quick BASIC on a Zenith 248 PC to convert ASCII files of C Cloud S coefficients listed above to Quick Basic BLOAD binary files. The program uses the actual ASCII names of the files from floppy disks to navigate the coefficients into the proper bins. Therefore, if new file names for coefficients on floppies are changed for any reason, the file names in DODAT must also be changed and the program recompiled. DODAT compresses each incoming coefficient in E15.7 format to integer form by multiplying it by 10000.0 and then storing the result as a 16 bit binary integer. (Interannual standard deviation coefficients are multiplied by 1000.0 before conversion to integer form.) Upon program completion, a list of maximum absolute values of coefficients encountered for each parameter is made available. This list was checked to insure that none of the integerized values exceeded 2**15th. The new compact coefficients are then suitably arranged for storage in the C Cloud S program as BLOAD files and could also be stored on a single floppy disk (360 version) along with other files.

REFERENCES

Boehm, A., 1992: CUB: A General Purpose Algorithm. Preprint of the Fifth International Meeting on Statistical Climatology, Toronto, Canada, American Meteorological Society, Boston, MA.

Burger, C. F., 1985: World Atlas of Total Sky Cover. AFGL-TR-85-0198 Air Force Geophysics Laboratory, Hanscom Air Force Base, MA, ADA-170474.

Dixon, W. J., chief editor, M.B. Brown, L. Engelman,M. A. Hill, R. I. Jennrich, 1988: BMDP Statistical Software Manual to accompany the 1988 software release. University of California Press, Berkeley, CA.

Hahn, C. J., 1987: Climatological Data for Clouds Over the Globe from Surface Observations. Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO.

Hwang, P. H. et, al., 1988a: "An Overview of the NIMBUS-7 THIR/TOMS Cloud Data Archive". Submitted for publication to the Bulletin of the American Meteorological Society.

Hwang, P. H., L.L. Stowe, H.Y. Michael Yeh, H. Lee Kyle, and the NIMBUS-7 Cloud Data Processing Team, 1988b: The NIMBUS-7 Global Cloud Climatology. Bull Amer. Meteor. Soc.,69 743-752.

"Reference Manual," 1977: USAFETAC DATSAV Data Base Handbook. USAFETAC-TN-77-2. Air Force Environmental Technical Applications Center, Asheville, NC.

"Reference Manual," 1986: DATSAV2 USAFETAC Surface Climate Database Users Handbook No. 4. USAFETAC/UH-86/004,USAF Environmental Technical Applications Center, Asheville, NC.

Smyth, A., J. H. Willand, and J. Steeves, 1991: Graphical Analysis of Bertoni's LOS Data. Prepared by ST Systems Corporation for Contract No. F19628-88-C-0089, Geophysics Directorate, Atmospheric Structures Branch, Hanscom AFB, MA, PL-TR-91-2172, ADA251592

Tompkins, C B and W.L. Wilson, Jr., 1969: Elementary Numerical Analysis Prentice-Hall Applied Mathematics Series, Prentice-Hall, Inc. Englewood Cliffs, N.J.

Warren, S. G., C. J. Hahn, J. London, R. M. Chevin, R. L. Jenne, 1986: Global Distribution of Total Cloud Cover and Cloud Type Amounts over Land. Prepared for United States Dept. of Energy, Office of Energy Research, Office of Basic Energy Sciences, Carbon Dioxide Research Division and National Center for Atmospheric Research, Boulder, CO. Technical Note NCAR/TN-273+STR (also DOE/ER/60085-H1).

Wellemeyer, C. G., 1986: NIMBUS-7 Observation Processing System (NOPS) Tape Specification #T399121 CMATRIX Tape. Prepared by SASC Technologies for the National Aeronautics and Space Administration Goddard Space Flight Center, Greenbelt, MD.

Westlake, J.R., 1968: A Handbook of Numerical Matrix Inversion and Solution of Linear Equations. John Wiley & Sons, Inc., New York, London, Sydney.

Willand J. H., 1988: Users Manual for C Cloud S Database. AFGL-TR-88-0060, Contract F19628-87-C-0046, Systems and Applied Sciences Corporation ADA-196497.

Willand, J. H.,and A. R. Boehm, 1988: "C Cloud S Climatology of Cloud Statistics." Proceedings of the Cloud Impacts on DOD Operations and Systems, 1988 Workshop (CIDOS-88).

Willand, J. H., and J. Steeves, 1991: Processing of Cloud Databases for the Development of an Automated Global Cloud Climatology. Interim Report. Prepared by HSTX for Contract No. F19628-88-C-0089 Geophysics Directorate, Atmospheric Structures Branch, Hanscom AFB, MA, PL-TR-91-2163, ADA251593.

APPENDIX A

```
      FUNCTION TETRA(A,B,C,D)
C     ***************************************************
C     ORIGINAL TETRA FUNCTION WAS WRITTEN IN BASIC
C     BY A. BOEHM OF ST SYSTEMS. THIS FORTRAN VERSION
C     WAS ADAPTED FOR THE CDC CYBER 60 BIT COMPUTER
C     BY J. H. WILLAND, ST SYSTEMS, ON 01/03/89.
C
C    PARAMETERS A, B, C, D MAY BE EITHER NORMALIZED ELEMENTS
C                               OR ACTUAL COUNTS.
C     RETURN THE TETRACHORIC CORRELATION.
C     REF. BOEHM 1976: TRANSNORMALIZED REGRESSION PROBABILITY,
C          AWSTR-75-259, P24.
C     ****************************************************
      DATA PI/3.1415926536/
C     --------------------------------------------------
C         Note that -999.0 is returned for robar if no solution.
      R=-999.0
      AD=SQRT(A*D)
      BC=SQRT(B*C)
      IF(AD.LE.0.0.OR.BC.LE.0.0)GO TO 99
C     -------------- FIRST GUESS --------------------
C     ----- PANOFSKY AND BRIER, SOME APPLICATIONS OF
C     STATISTICS TO METEOROLOGY, 1965, PP. 103-104.
      RSIN=SIN(PI/2.*((AD-BC)/(AD+BC)))
      R=RSIN
      XN=A+B+C+D
      P1=(A+C)/XN
      HQ=-ENORM(P1)
      P2=(A+B)/XN
      QK=-ENORM(P2)
      AN=A/XN
      R1=0.
      D1=P1*P2-AN
C     --------------  ITERATIVE SOLUTION  -------------
      DO 10 K=1,25
        D0=FLNORM(HQ,QK,R)-AN
        IF(ABS(D0).LT.1.0E-7)GO TO 99
        RS=R
        IF(ABS(D0-D1).LT.1.0E-7)GOTO 99
        R=R-D0*(R1-R)/(D1-D0)
        D1=D0
        R1=RS
c          Next line was corrected 12/2/92. (1.,R) was (R,R). J. W.
        IF(ABS(R).GE.1.0)R=(SIGN(1.,R)+RS)/2.
   10 CONTINUE
   99 TETRA=R
      RETURN
      END
```

```
      FUNCTION TFUNC2(A,B,C)
C     ************************
C     AVOIDS DIVISION BY ZERO.
C     ************************
      IF(ABS(C).LT.1.0E-8)THEN
        Y=SIGN(B,B)*(1.0-PNORM(ABS(A)))/2.
        TFUNC2=Y
      ELSE
        TFUNC2=TFUNC(A,B/C)
      ENDIF
      RETURN
      END
```

```
      FUNCTION TFUNC(HQ,A)
      DATA PI/3.1415926536/
C     ************************
C     BINORMAL SECTOR INTEGRAL.
C     REF. YAMAUTI,(1972): STAT. TABLES AND FORMULAS WITH COMPUTER
C                      APPLICATIONS, JAP STAT. ASSC.,PROGRAM 15.
C     ************************
      IF(ABS(HQ).LE.1.0E-50)GO TO 20
      A1=ABS(A)
      H1=HQ
      IF(A1.LT.1.0)GO TO 9
      H1=A1*HQ
      A1=1./A1
    9 AA=A1*A1
      HH=H1*H1/2.
      DHH=ALOG(HH)
      SUM2=EXP(-HH)
      SUM1=1.-SUM2
      FACT=0.0
      S=-AA
      DO 10 J=1,80
      X=J
      FACT=FACT-ALOG(X)+DHH
      SUM2=SUM2+EXP(-HH+FACT)
      C=S*(1.-SUM2)/(2.*X+1.)
      SUM1=SUM1+C
      IF(ABS(C).LT.1.0E-10)GO TO 11
      S=-S*AA
   10 CONTINUE
      PRINT 100,HQ,A
  100 FORMAT(" TFUNC HQ A",2F10.4,1X,"DID NOT CONVERGE.")
   11 T=(ATAN(A1)-SUM1*A1)/(2.*PI)
      IF(ABS(A).LE.1.0)GO TO 19
        AA=PNORM(-HQ)
        HH=PNORM(-H1)
         T=(AA+HH)*.5-AA*HH-T
   19 IF(A.LT.0.0)T=-T
      TFUNC=T
      GO TO 99
   20 TFUNC=ATAN(A)/(2.*PI)
   99 RETURN
      END
      FUNCTION FLNORM(U,V,R)
C     ******************************************
C     COMPUTE THE INTEGRAL OF THE BIVARIATE NORMAL
C     DISTRIBUTION UP TO LIMITS U AND V.
C     REF OWEN,1980: A TABLE OF NORMAL INTEGRALS,
C         COMMON STATISTICS.-SIMULA-COMPUTA.,
C         B9(4), 389-419, EQ 3.1.
      DATA PI/3.1415926536/
C     ------------------------------------------
      IF(ABS(U).LT.1.0E-8 .AND. ABS(V).LT.1.0E-8)GO TO 999
      H=-U
      XK=-V
      Y=(PNORM(H)+PNORM(XK))/2.
      IF(H.LT.0..AND.XK.GE.0.)Y=Y-.5
      IF(XK.LT.0..AND.H.GE.0.)Y=Y-.5
      Y=Y-TFUNC2(H,XK-R*H,H*SQRT(1.-R*R))-
     X   TFUNC2(XK,H-R*XK,XK*SQRT(1.-R*R))
      FLNORM=Y
      RETURN
  999 FLNORM=.25+ASIN(R)/(2.*PI)
      RETURN
      END
```

```
      FUNCTION PNORM(E)
C     ************************
C     END TO PROBABILITY
C     ADAPTED FOR THE CDC 60 BIT COMPUTER BY J. WILLAND (STX)
C     CUMULATIVE NORMAL.
C     INTEGRAL OF THE STANDARD NORMAL DISTRIBUTION
C     FROM INFINITY TO E.
C     REF. ABRAMOWITZ AND STEGUN  (1964); HANDBOOK OF
C          MATH. FUNCTIONS, NBS, EQ. 26.2.19.
C     *************************
      A=ABS(E)
      A=((((((5.621326E-6*A+5.105517E-5)*A+3.968616E-5)*A+3.422739E-3)
     X*A+2.2077E-2)*A+5.207516E-2)*A+1.)4.427378E-2)
      A=A**(-16)
      IF(E.GT.0.0)A=1.-A
      PNORM=A
      RETURN
      END
      FUNCTION ENORM(P)
C     *************************
C     PROBABILITY TO EQUIVALENT NORMAL DEVIATE (END)
C     ADAPTED TO THE CDC 60 BIT COMPUTER BY J. WILLAND (STX)
C     ALGORITHM AS111, APP STAT 26 1977  P118.
C      INVERSE OF STANDARD NORMAL DISTRIBUTION
C     INTEGRATION FOUND IN PNORM.
C     ************************
      Q=P-.5
      IF(ABS(Q).GT.0.42)GOTO 10
        R=Q*Q
        R=Q*(((-25.44106*R+41.3912)*R-18.615)*R+2.506628)/
     X  ((((3.130829*R-21.06224)*R+23.08337)*R-8.473511)*R+1.)
      ENORM=R
      RETURN
   10   IF(Q.LE.0.0)THEN
        R=P
      ELSE
        R=1.-P
      ENDIF
      IF(R.LT.1.0E-500)R=1.0E-500
      R=SQRT(-ALOG(R))
      R=(((2.321213*R+4.850141)*R-2.297965)*R-2.787189)/
     X  ((1.637068*R+3.543889)*R+1.)
      IF(Q.LT.0.0)THEN
        ENORM=-R
      ELSE
        ENORM=R
      ENDIF
      RETURN
      END
```

```
      SUBROUTINE SKYRO(R,RBAR)
C     **********************************
C
C     MEAN CORRELATION GIVEN SCALE DIST. BY A. BOEHM AND J. WILLAND
C     ST SYSTEMS 109 MASS.AV, LEXINGTON MA.
C    VERSION 3.0 2/19/91
C    ASSUMES 2424 KM SKY DOME AND WAVELENGTH=256R THAT WAS USED IN BAA
C    EDGE IS DISTANCE(IN WAVELENGTHS) ALONG EDGE OF SQUARE WITH 2424 KM AREA
C          R=SCALE DISTANCE (KM)
C       RBAR=RETURN AVERAGE CORRELATION (ROBAR)
C
C     **********************************
      DATA N/21/
C     EDGE=SQRT(2424.)/(256.*R)
      EDGE=.19232/R
C     SUM=(21*21)/4 = 110.25
      SUM=110.25
      DX=EDGE/(N-1)
        DO 20 I=1,N-1
          DO 10 J=0,N-1
          SIG=SQRT(FLOAT(I)**2+FLOAT(J)**2)*DX
          D=SIG - INT(SIG)
          C=1.-3*D +2.*D*D
          IF(SIG.GT.1.) C = D*C/SIG
          SUM=SUM + (N-I)*(N-J)*C
   10 CONTINUE
   20 CONTINUE
      RBAR=4.*SUM/N**4
      RETURN
      END
```